Vol. 68, No. 4 July 1961

Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

CONTENTS

PUBLISHED BIMONTHLY BY THE

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals. However, reports of original research may be included in a major theoretical article which attempts to integrate several related original studies.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted.

All manuscripts must be submitted in duplicate.

Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript, in case there is loss in the mail. Manuscripts should be addressed to the Editor, Richard L. Solomon, Psychology Department, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania.

Arthur C. Hoffman
Managing Editor

Helen Orr
Promotion Manager

Published Bimonthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1333 Sixteenth St. N.W., Washington 6, D. C.

$10.00 volume $2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

VOL. 68, NO. 4 JULY 1961

PSYCHOLOGICAL REVIEW

TEN YEARS OF MASSED PRACTICE ON DISTRIBUTED PRACTICE [1]

BENTON J. UNDERWOOD

Northwestern University

For the past 10 years a series of studies dealing with the influence of distributed practice on verbal learning has been carried out in the Northwestern University laboratories. The purpose of the present paper is twofold: to present the critical variables which have emerged from this series of studies, and to suggest certain conceptual notions which at the present time seem useful in trying to understand the effects of distributed practice and related phenomena.

The primary empirical goal at the time this series of studies was initiated was a straightforward one, namely, to determine the range of conditions and materials within which distributed practice facilitated learning or retention. The fact that 10 years have passed since this goal was established indicates that it has proven to be an elusive objective to attain. Indeed, no implication should be drawn from the present paper that the goal has now been reached; the pursuit continues. Therefore, the "critical variables" to be presented, while representing a considerable experimental distillation over the years, may yet change with continuing work and new perspectives.

To say that certain variables are more critical than others necessarily implies some evaluation. For the phenomenon being dealt with in this paper, the evaluation is based on three criteria. First, and fundamentally, the positive effects of distributed practice are not found unless a critical variable is involved. Second, the effects of other variables can be shown to be a consequence of the simultaneous change in a critical variable. Finally, using the particular critical variables chosen, conceptual notions are suggested which in turn lead to further testable propositions.

There should be no misunderstanding about the conceptual notions to be presented in this paper. They are a posteriori notions; the sifting of the facts of massed versus distributed practice led to an assertion about critical

[1] The substance of this paper constituted the Presidential Address given before Division 3, American Psychological Association, Chicago, 1960.

Most of the work reported here was supported by Contracts N7onr-45008 and Nonr-1228(15) between Northwestern University and the Office of Naval Research. Over the 10 years of this work five extraordinarily capable research associates have been involved: Ross L. Morgan, E. James Archer, Jack Richardson, Rudolph W. Schulz, and Geoffrey Keppel. The author is deeply indebted to them all. For certain analyses and data first presented here, the aid of Schulz and Keppel has been invaluable, and the manuscript has benefited by their critical readings.

variables, and then it was found possible to cast these variables into a theoretical scheme which had some generality. The theory was not originated as a deductive one, but neither was it ad hoc. Its usefulness, no matter how momentary, can be justified on the grounds that it makes some sense out of a number of phenomena and, in addition, yields a number of testable propositions.

Because the theory has emerged from the various analyses leading to the selection of the critical variables, the order of the argument in this paper will maintain some historical integrity as follows: the nature of the data leading to the selection of the critical variables, the theoretical notions by which these critical variables are asserted to produce the phenomena associated with the massed-distributed problem, and a consideration of other phenomena which may be incorporated within the system. However, before initiating the central arguments, it is necessary to dispose of certain preliminary matters.

The Basic Operations. The basic procedures used in the Northwestern studies, procedures which are quite comparable to those used by other investigators, are as follows. A trial is defined as the single presentation of all items in a list of verbal units. The list may be presented for a constant number of trials or until the subject achieves some arbitrary level of performance, such as one perfect recitation. The central variable is the length of the interval between trials. If the interval is short, say 2 seconds to 8 seconds, learning is said to be by massed practice (MP). If the interval is longer—15 seconds or more—learning is said to be by distributed practice (DP). In actual practice the DP interval has not exceeded 3 minutes. The simple empirical problem is to determine under what conditions learning under DP differs from learning under MP. Several studies have also dealt with the effect of MP and DP on long-term retention. However, because of the scope of the material to be covered on the effects of MP and DP on learning, their effects on retention will not be covered in this paper.

Magnitude of the Effects. When DP is introduced in the acquisition of a motor task, such as a pursuit rotor, enormous facilitative effects on performance are observed. There is no comparison to be made between the magnitude of such effects and those which occur as a consequence of DP in verbal learning. Facilitation by DP in verbal learning occurs only under a highly specific set of conditions, and the magnitude of the effect when it does occur is relatively small. The reliability of certain DP effects must, in some cases, be based upon consistency from experiment to experiment rather than upon a strict statistical criterion for any given experiment. If one wishes to use an efficiency measure for learning, it would be very inefficient to learn by DP; the subject would be much further ahead to learn by MP if total time to learn (including the rest intervals in DP) is the criterion. Even under the most favorable conditions for facilitation by DP, one could not recommend its use in an applied setting where verbal materials are to be mastered.

First Critical Variable: Response-Term Interference

In this section evidence will be summarized which has led to the conclusion that a certain amount of response-term interference must be present before DP will facilitate acquisition. However, some background preparation is necessary.

Two-Stage Conception

In recent years it has been found useful to view verbal learning as occurring in two stages. The first stage is the response-learning stage during which the subject must acquire the responses as responses; they must become readily recallable units. The second stage is the associative stage during which the response must be attached or hooked up to a specific stimulus. A more complete elaboration of these two stages has been given elsewhere (Underwood & Schulz, 1960); not only do these stages seem to be logically necessary but the behavior of subjects in mastering lists appears to correspond closely to the two stages. Furthermore, variables may have different effects during the two stages. For example, high similarity among responses in a paired-associate list facilitates response learning but retards associative learning (Underwood, Runquist, & Schulz, 1959).

The relevance of the two-stage conception for understanding the effects of DP is this: when DP facilitates verbal learning the evidence points strongly toward the fact that it does so because of interference operating during the response-learning stage. This would *exclude* interference in attaching responses to particular stimuli. It would *include* any interference which impedes the learning of responses as responses. This is most clearly seen in learning low-meaningful units, such as consonant syllables. The interference may prevent the ready learning of the correct sequences of the letters of the consonant syllables. Or, to say this another way, the interference may retard response integration.

The evidence leading to this conclusion will now be given in summary form.

1. In a recent study (Underwood & Schulz, 1961a), it was shown that if interference was built up among responses across paired-associate lists, DP facilitated learning. However, if interference was built up among stimuli across lists (with minimum interference in acquiring responses), DP actually retarded learning.

2. In a study of verbal-discrimination learning and DP (Underwood & Archer, 1955), intralist similarity and rate of presentation of the pairs was varied. Under conditions which would normally produce facilitation by DP in serial or paired-associate learning, no effect was found. In verbal-discrimination learning the subject does not have to acquire the response terms in the sense that he does in serial or paired-associate learning. Rather, he is presented a pair of items and he merely has to indicate which member of the pair is correct; no response integration is required.

3. In a study of concept recognition (Underwood, 1957a) where the response was a common word, no facilitation by DP was found in spite of the fact that heavy interference was produced because of the overlap among concept stimuli.

4. Paired-associate lists of 16 pairs were constructed in which interference was varied in amount by using very familiar words representing the same concept (Underwood & Schulz, 1961b). Although widely different rates of learning were observed as a consequence of the differential interference, DP did not facilitate learning at any level of interference. In this task there was no problem of integrating responses; they were common words which were quickly acquired as responses. The interference occurred in attaching them to particular stimuli.

Such evidence as the above has led to the conclusion that DP will facilitate

acquisition only when the interference impedes the integration of the response as such. In most of the situations occurring in our studies this has involved the integration of letters forming verbal units which are not words. However, an argument will be advanced later that if interference occurs among syllables in words, DP may also facilitate acquisition.

Types of Interference in Response Integration

The available evidence indicates that as long as interference attains a certain level in the response-acquisition stage, the source of the interference is irrelevant. The interference may derive from formal similarity by duplicating letters among nonsense and consonant syllables (e.g., Underwood & Schulz, 1959). It may stem from habits developed in learning previous lists in the laboratory (e.g., Underwood & Schulz, 1961a). Finally, the interference may come from letter sequence habits which have been built up during the lifetime of the subject. Thus, if the response required of the subject involves sequences of letters which run counter to strongly established letter-sequence habits, DP may facilitate acquisition (Underwood & Schulz, 1961b).

Analytical Implications

The evidence summarized above strongly suggests that interference in response integration is responsible for facilitation by DP. Furthermore, it has been shown that similarity among stimulus terms may actually retard learning by DP. Certain implications for analysis follows from this state of affairs.

1. In serial learning the functional stimulus for any given item in the list is essentially unknown. It may be serial position, the immediately preceding item, several preceding items, or some complex of all of these. Therefore, serial learning is not a task providing sufficient isolation between stimulus and response functions to produce critical theoretical decisions. This does not mean, of course, that serial learning will not respond to DP. Indeed, it has been found generally "easier" to get positive effects of DP for serial than for paired-associate lists, probably because of differences in rate of presentation usually employed (Hovland, 1949). But, if the inferences made earlier are correct, when the investigator cannot specify the locus of interference (among stimulus terms or among response terms), the worth of a task is much reduced for the study of DP.

2. The paired-associate task is most suited for decisive testing of hypotheses about how DP produces its effect. However, if interference occurs among stimulus terms, among response terms, and among stimulus and response terms—all simultaneously—the analytical capabilities of the task are greatly reduced. The use of counterbalanced designs may well "throw in" interference of an unspecifiable nature across lists. Thus, when such designs are used and when interference is ostensibly independently manipulated among stimulus terms and among response terms within lists, interlist interference may completely eliminate the analytical usefulness of these studies; that this seems to have been the case has been detailed elsewhere (Underwood, 1954).

SECOND CRITICAL VARIABLE: LENGTH OF INTERTRIAL INTERVAL

The length of intertrial interval is, obviously, a critical variable since it is an integral part of the defining operations for the phenomenon under consideration. So, therefore, the question

asked about this variable concerns phenomena which may be correlated with variations in the intertrial interval.

If one asks the beginning student in elementary psychology what would happen if a rest interval were introduced as he was endeavoring to learn a list of words, a likely response would be that he would forget some of what he had learned. The logic of this position seems unassailable if we use rest intervals of an hour, a day, or a year. The line of evidence to be pursued is that measurable forgetting occurs over the short intervals used to define DP.

1. A study has been reported (Underwood & Schulz, 1961a) in which interference among stimuli in paired-associate lists was built up across four lists. In learning this fourth list, DP (1-minute interval) was somewhat inferior to MP. Since the responses in these lists for both studies were adjectives of low intralist and interlist similarity, the effect observed appears to result from interference in the associative or hook-up stage of learning. It is presumed that the DP interval results in more proactive interference than the MP interval. The effect of this interference on short-term forgetting may be indexed by determining on which trial responses were first correctly paired with their stimuli. A tally was made of the trial on which each response was first correctly paired with its stimulus (combining all subjects). Then, the percentage of total items first given correctly on each trial was determined; for presentation here, blocks of two trials have been combined. As seen in Figure 1, a greater percentage of responses was first paired correctly on

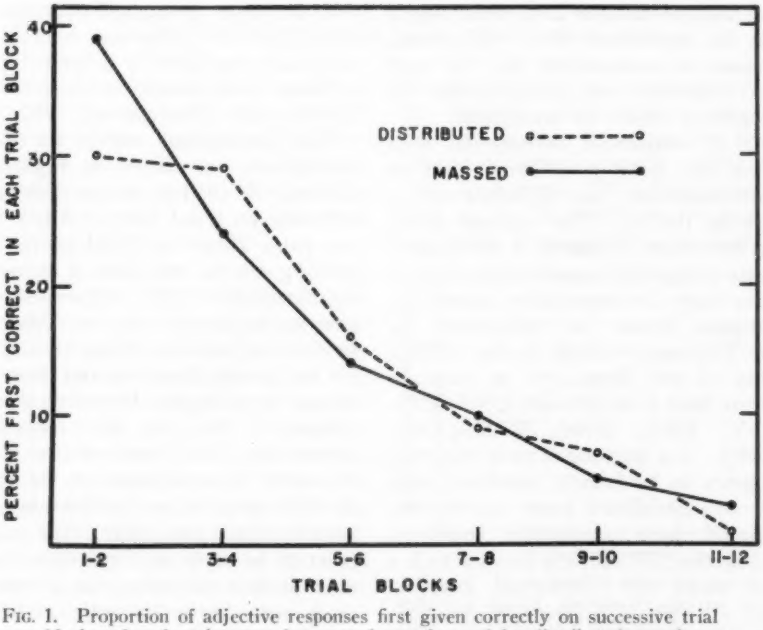

FIG. 1. Proportion of adjective responses first given correctly on successive trial blocks when learning was by massed practice and by distributed practice.

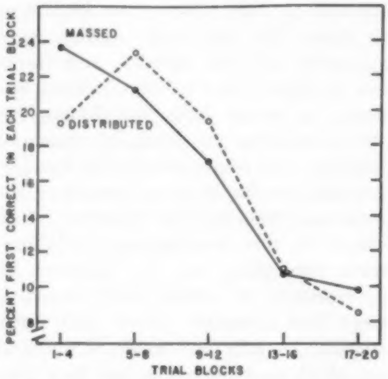

FIG. 2. Proportion of trigram responses first given correctly on successive trial blocks when learning was by massed practice and by distributed practice.

Trials 1 and 2 when learning was by MP than when it was by DP. There were 48 subjects in each condition; the t between the mean number of different items correctly anticipated on the first two trials is 2.03, which meets the 5% significance level. Of course, it must be remembered that for such short intervals no great amount of forgetting would be anticipated.

2. A comparable analysis has been made for quite a different kind of paired-associate list (Underwood & Schulz, 1961b). The response terms in these lists consisted of three-letter units (trigrams) constructed so as to have very low-associative connection between letters as determined by the Underwood-Schulz tables (1960). One of the three sets of response terms used is as follows: DSU, RZL, CFY, XBN, IGW, TPM, OVJ, KHQ. The learning of such responses appears to be heavily interfered with by well-established letter associations. Stimuli were single-digit numbers. Altogether, 70 subjects learned such a list under DP (30-second interval) and 70 with MP (4-second interval) for 20 trials. The learning under DP was not appreciably better than under MP (a matter for later discussion). The percentage of responses first given correctly in successive blocks of four trials is shown in Figure 2. This figure indicates that under DP the responses on the average are first given correctly at a later point in learning than under MP. The mean difference between MP and DP on the first block of trials fails to achieve an acceptable level of statistical significance, but that a true difference in forgetting is occurring is suggested by the next set of data.

3. A list was made up of single-digit numbers as stimuli and the eight above listed trigrams as responses. Thirty-six subjects, naive to verbal-learning experiments, were assigned on a random basis to one of two groups. Each group received 20 acquisition trials. One of the groups was given a 2-minute rest after Trial 9, and the other group a 2-minute rest after Trial 18. The rest interval in each case was filled by a symbol cancellation task commonly used to fill DP intervals (Underwood, 1952a).

The performance curves for these two groups are shown in Figure 3. Not only do the two groups differ significantly on Trial 10 ($t = 4.42$), but loss from Trial 9 to Trial 10 for the group given the rest interval is highly significant ($t = 3.05$). However, differences in performance on Trial 19, or the loss between Trials 18 and 19 for the group given the rest interval, do not even approach statistical significance. The fact that forgetting occurs after Trial 9 and not after Trial 18 would be anticipated on the basis of differences in level of learning attained at those two points. The major point to be made by these data, however, is that forgetting can be measured over short intervals early in learning when the response terms are

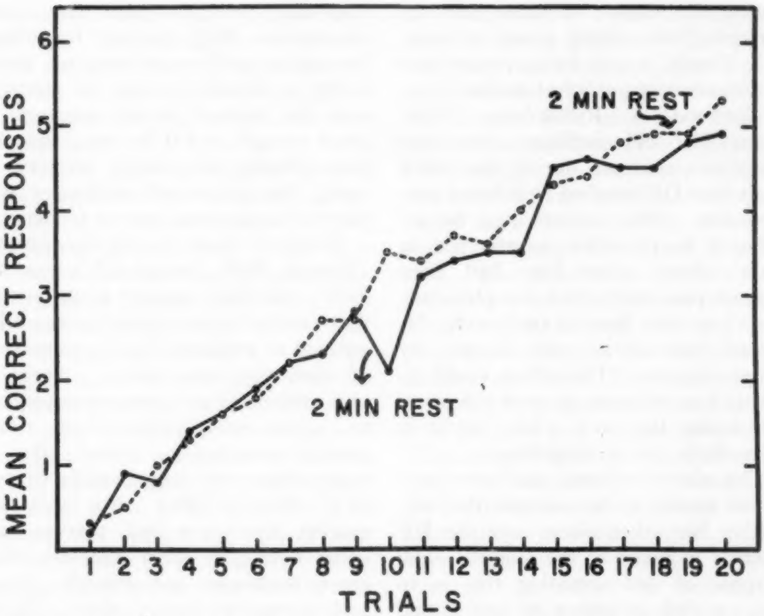

FIG. 3. Forgetting of trigram lists over a 2-minute interval when the interval is introduced after the ninth trial and after the eighteenth trial.

difficult ones to learn (presumably because of interference from stronger letter-association habits). That forgetting may occur over a 2-minute interval is quite in line with previous results using serial lists of consonant syllables (Underwood, 1957b).

4. In a published study (Underwood & Schulz, 1959) data have been presented on the learning of a serial list with high formal intralist similarity among consonant syllables. Four different intertrial intervals were used, namely, 2, 8, 17, and 38 seconds. Over an interval the subject may forget the response as such, or he may forget its proper position in the serial list. There is no evidence in this experiment that the syllables as such were forgotten over these short intervals, but the initial retention of the proper positions of the items in the list was impaired as a consequence of the interval. The measure used to demonstrate this is simply the mean number of items incorrectly positioned when first given. This measure is related to the length of the interval in Figure 4. The longer the interval the

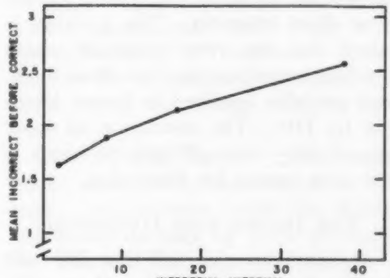

FIG. 4. Mean number of serial consonant syllables placed incorrectly before being placed correctly as a function of the length of the intertrial interval.

greater the number of items given incorrectly before being given correctly.

5. Finally, it may be mentioned that in a number of published studies (e.g., Underwood & Richardson, 1958) where DP did facilitate acquisition there was evidence during the initial trials that DP resulted in inferior performance. This effect may be attributed to proactive interference in cases where other lists had been learned previously. But the phenomenon has also been found with 38-second intervals in serial learning by naive subjects. The effect could be due to loss of warm-up over the interval during the early trials, but it is more likely due to forgetting.

The above evidence has been presented merely to demonstrate that forgetting has taken place over the DP interval in many of the studies. The purpose of demonstrating this is to suggest that in trying to understand how DP produces a positive effect the understanding of processes producing forgetting must be given a central role. To attain this goal, the evidence presented above is no substitute for extensive systematic studies of short-term retention using different materials for which interference possibilities are known. Since such studies are not available, certain assumptions will have to be made about how forgetting occurs over short intervals. The position is taken that the very processes which produce forgetting over the short intervals are also involved in better learning by DP. The resolution of these superficially contradictory positions is the next matter for discussion.

The Interaction Hypothesis

It has been asserted that DP will facilitate acquisition only when some minimal level of interference is present in the response-acquisition phase of learning. An interference theory of forgetting presumes that this same interference will produce forgetting. Forgetting will occur because interfering tendencies recover in strength over the interval; if the recovery is great enough, and if the intial association defining the correct response is weak, the recovered tendencies will block or replace the correct tendencies.

Evidence from recent experiments (Barnes, 1960; Barnes & Underwood, 1959) give firm support to the notion that verbal associations are extinguished or weakened by the occurrence of conflicting associations. There is also evidence that a process analogous to spontaneous recovery of extinguished associations occurs; this is fairly clear over long periods of rest (e.g., Briggs, 1954). But it is also evident that some such process may occur over very short intervals when error tendencies are initially strong and correct tendencies weak. In almost all of our studies of DP it has been found that more overt errors are made early in learning under DP than are made under MP. The one assumption which must be made, and it is an assumption for which direct independent evidence in the learning of verbal units is lacking, is that extinction of associations following recovery attains a greater "depth" than when such recovery does not occur. Under highly MP conditions such recovery is presumed not to occur (at least to any extent). Under DP there is successive recovery of error tendencies and successive extinctions. Under MP the error tendencies essentially remain suppressed but, at the same time, reduce the effective strength of the correct response tendency. Early in learning this reduction in effective strength of the correct response tendency may be less than that produced by the recovered tendency in DP; hence, forgetting under DP may be shown to be

greater than under MP over the short intervals. But, as interfering associations are weakened by successive extinctions in DP—and not in MP—the situation becomes reversed as trials continue.

With the above conception in mind it can be seen why obtaining facilitation by DP is dependent upon a highly specific set of conditions. If DP is to produce facilitation, the following conditions must be met: (*a*) there must be some minimal amount of interference in response acquisition so that the integration of the components of the correct response by the subject is slow, (*b*) error tendencies must recover enough so that successive extinctions can be effective, but (*c*) the recovery must not be so great as to block or replace the correct association over several trials. The situation is most delicate early in learning when correct response tendencies are relatively weak and error tendencies relatively strong. Thus, the two critical manipulable variables are amount of interference and length of DP interval because these two variables allow indirect manipulation of strength of error tendencies and amount of recovery of error tendencies.

One of the implications of the above conception is that given at least some minimal amount of interference in response integration, a DP interval could be found which would facilitate acquisition. Data already published and data to be published do not completely conform to this expectation so that considerable evaluation is necessary.

Length of Interval

An examination of studies in which length of DP interval is varied does not give much support to the idea that length of DP interval is critical, given a certain level of intial interference. This is true in our studies and also in those of other investigators (e.g., Wright & Taylor, 1949). However, note can be taken of certain studies where there is suggestive evidence that longer DP intervals result in poorer learning than do shorter ones. For example, in one study (Underwood, 1953c) two independent experiments were done in which 60-second and 120-second DP intervals were used. The performance in both studies was poorer with 120 seconds than it was with 60 seconds. Such a result also occurred in another study where the intervals were 30 seconds and 60 seconds (Underwood, 1952a), but not in another of near comparable design (Underwood, 1952b). In still another study (Underwood & Richardson, 1955) there was no inversion in the relationship between length of DP interval and rate of learning up to 3 minutes. However, the present belief is that this relationship was confounded by a subject selection. The learning of difficult lists was carried until the subject achieved a criterion of one perfect trial. Total learning time was limited to a 50-minute period. If the subject did not achieve the criterion within that period he was dropped. The number of subjects dropped was directly related to length of DP interval. It is possible, therefore, that the longer the interval the greater the selection of subjects who could learn under the long DP intervals. It may be that those selected represent subjects whose short term memory was good.

The Wright-Taylor study, noted above, used DP intervals up to 8 minutes. Performance under the 8-minute condition was as good as under the shortest DP interval, which was 75 seconds, and all DP conditions were superior to MP (2 seconds). There are possibilities of rationalizing the results of this study so that it does not

stand in such apparent contradiction to the position stated earlier. Interference involved would be fairly low and entirely of an intralist nature, since the subjects were apparently naive. It is possible that all recovery of intralist error tendencies that would occur had taken place in the shortest DP interval (75 seconds). Forgetting which might occur because of interference from outside sources would be slight for these naive subjects over the longest DP interval (8 minutes). Finally, cartoon reading was used to fill the DP intervals; this may be quite an ineffective method to prevent rehearsal.

In short, there is no evidence available which denies the possibility of full exploration of the notion that the critical interaction involved is that between amount of interference and length of DP interval.

Degree of Response Integration

Three paired-associate lists were made up in which response terms were trigrams of three different degrees of integration (Underwood & Schulz, 1961b). The list of initially low integration was made up of items comparable to those presented earlier in this paper. The higher the degree of integration, the easier it is to learn the list. A DP interval of 30 seconds did not produce positive effects for any of the three lists. However, the list of poorly integrated trigrams showed evidence of forgetting (as indicated in Figure 2) with an indication that DP was producing some facilitation near the end of the 20 trials given. This suggests that a shorter DP interval, perhaps 15 seconds, would result in significant facilitation in learning. If this is the case, however, why did the list of somewhat higher integration not show facilitation over 30 seconds? There are two possibilities. This list was much easier to learn than the list of low integration. It may not have met the minimum requirements for amount of response interference. If a list were constructed which is in between the low and medium list in terms of degree of integration such a list might show facilitation with a 30-second DP interval. The other possibility is that the relatively weak error tendencies in the medium list require a longer DP interval for recovery, such recovery being necessary for DP to facilitate.

It can be seen that in making the above predictions it is again necessary to specify a highly particular set of conditions before DP will facilitate learning. One strong reason for confidence in such predictions lies in the results of the following experiment, an experiment which was viewed as a method of manipulating initial response interference (Underwood & Schulz, 1961b).

Two paired-associate lists in which two-letter units were the response terms were constructed. In one list each of these bigram responses had low integration between the two letters, and in the other the integration was high. The units having low integration were: EY, CF, XV, DS, QW, JH, MK, RZ. Those having high integration were: MN, RA, CO, QU, XY, ET, JK, DI. The stimuli for both lists were single-digit numbers. The list of low initial integration was much more difficult to learn than the one with high integration. Furthermore, DP (30 seconds) significantly facilitated the learning of the list with poorly integrated items and had no effect on the list in which the bigrams were well integrated.

Thus, with a 30-second DP interval, the learning of the lists in which trigrams were responses was not facilitated, but the learning of bigrams was facilitated. That the number of letters

is not the critical variable is shown by the fact that with single letters as responses DP did not facilitate (Underwood & Schulz, 1961b). The conclusion from these experiments is that there is a very sharp interaction between length of the DP interval and the amount of initial response interference. With the poorly integrated bigrams, the 30-second interval may have been about optimal; with the poorly integrated trigrams, this interval may have been too long. It is also possible that if response interference is too high, DP will never facilitate acquisition if the DP intervals are introduced at the start of learning. That is to say, any recovery of error tendencies may so retard learning that MP will be superior. However, this is not expected to be the case; rather, when interference is heavy, the expectation is that the DP interval must be short in order for facilitation to occur.

Words

To maintain that DP facilitates acquisition only when interference occurs in acquiring the responses as such requires a consideration of instances in which words have been used and in which DP has facilitated acquisition. The problem provided by these instances is simply this: if the interference must obtain during the response-integration phase, why should words ever be facilitated by DP? Before considering possible answers to this question, the facts concerning the effects of DP in the learning of word lists need to be reviewed.

Serial lists of adjectives, with relatively low intralist meaningful similarity, have been shown to be learned faster under DP than under MP (Underwood, 1951, 1953b; Underwood & Goad, 1951; Wilson, 1948). The differences have not always been statistically significant, but the consistency argues that DP is having some effect. In one study (Underwood & Goad, 1951) in which all adjectives within the serial list had the same core of meaning the effects of DP were greater than if the adjectives had relatively low meaningful similarity. In another study (Underwood, 1953b), where similarity was manipulated among clusters of words within serial lists, the effects of DP were small and did not differ as a function of the degree of similarity. No effect of DP has been noted when paired-associate lists of adjectives have been used (Underwood, 1953a).

All of the above studies used two-syllable adjectives. The position is taken here that difficulties may occur in integrating two-syllable adjectives in the same sense that these difficulties may occur in integrating two letters within a nonsense syllable or a consonant syllable. For example, consider a list in which two of the words are ROUNDED and CRUMBLING. Interference in integration could be indexed by the subject responding with ROUNDING and CRUMBLED. Thus, when syllables are interchangeable and the interchange still makes words—whether of the same part of speech or not—interference in response integration may occur. Such merger responses do in fact occur in the record sheets.

It would be presumed that when interference occurs among syllables of the words it would occur most severely when the correct syllable sequence has relatively weak associative connection and the incorrect relatively high. The situation may be viewed in exactly the same way as was done previously for letter sequences. It would be possible to construct lists in which the likelihood for interference varied from low to high among the syllables making up the words within a list. So far as is known, however, no such system-

atic data are available. Therefore, only the assertion can be made that syllabic interference is responsible for instances in which DP has facilitated the learning of word lists. Two sets of data are available, however, which indirectly support the position. These data have not been published elsewhere.

It was noted above that in one study (Underwood & Goad, 1951), learning of serial lists was facilitated by DP when all items in the list had the same core of meaning. In this counterbalanced study subjects learned a total of eight lists so that interlist syllabic interference could be built up. If the source of the interference is largely of an interlist nature, DP should not facilitate the acquisition of these lists if the subjects were relatively naive. A test of this implication was made. Two of the original high-similarity lists were used, each being learned by 12 subjects under MP (2 seconds) and 12 subjects under DP (30 seconds). Thus, 24 subjects learned under DP and 24 under MP. The lists were presented for 20 acquisition trials. While learning was more rapid under DP than under MP, the difference was far from being significant statistically. This finding, therefore, argues that the original results were at least in part an interlist effect; since several lists had been learned the possibilities of syllable interchange was much greater than in the present study.

In another study a serial list was constructed of 14 words. The 14 words, seven pairs of highly associated words from the Minnesota free-association norms (Russell & Jenkins, 1954), were as follows: TABLE, CHAIR, BED, SLEEP, NAIL, HAMMER, GIRL, BOY, SCISSORS, CUT, LOW, HIGH, LONG, SHORT. The words were randomized in terms of serial position within the list presented to the subject. The notion was that such a list would have rather high internal interference because of the strong associative connections existing between pairs of words. However, there should be low syllabic interference because many of the words have only one syllable. Response learning should take place very quickly but the existing associative connections should retard the associative phase of learning. Thirty naive subjects were presented the list (for 20 trials) under MP and 30 subjects under DP (2 seconds and 30 seconds, respectively). Learning by MP was not superior to that by DP. Thus, a word list with rather heavy interference, but an interference that did not prevail among syllables, was not facilitated by a 30-second DP interval.

It should be clear that the above two studies only indirectly support the idea that syllabic interference of a certain amount must be present before DP will facilitate the learning of word lists. Only studies specifically designed to vary syllabic interference will give a decision as to whether or not the position taken here is appropriate.

RELATED VARIABLES AND PHENOMENA

Serial Position Curve

In a recent study (Underwood & Schulz, 1959) it was noted that when DP facilitated the acquisition of serial lists the shape of the serial position curve differed for MP and DP. Learning under MP gave the classical bowed curve, clearly skewed so that the point of maximal difficulty was past the center of the list, and with the last few items in the list clearly being learned more slowly than the first few items. Under DP, however, the curve was almost symmetrical: very little skew was evident and the last items

in the list were learned almost as rapidly as the first items in the list. The difference in the shapes of these curves led to the conclusion that DP facilitated learning in a direct relationship to the position of the item in the serial list: those items near the end of the list were greatly facilitated, those near the beginning were facilitated very little. The question raised here is whether or not this phenomenon should be given any theoretical importance in trying to understand how DP produces its effect. The conclusion is that it should not be given such status. The reason for this is that it is a phenomenon limited entirely to subjects naive to serial learning as experienced in the laboratory.

An examination has been made of a number of our previous studies in which subjects served for several days and learned several lists (e.g., Underwood & Richardson, 1955). In every case it was found that in learning the first list the above described difference in the shape of the serial position curve as a function of MP vs. DP was found, but after the subjects were practiced the differences disappeared and both the MP and DP serial position curves assumed the shape of the classical curve. This was true in spite of the fact that DP had facilitated learning at all stages of practice. Furthermore, with naive subjects learning lists which were not facilitated by DP, the last few items in the list were learned more rapidly under DP than under MP but at the expense of inferior performance on other items in the list, primarily those at the beginning.

Because of these facts it seems most reasonable to interpret this phenomenon as a side effect resulting from the introduction of the DP intervals to naive subjects. It may represent a short-term memory effect based on recency, or it may be that naive subjects adopt a different strategy of learning when given DP than when given MP. In any event, at the present time there seems to be no necessity to attach theoretical importance to the phenomenon when the theoretical interest is in understanding how DP produces its facilitating effect on learning.

Rate of Presentation

Previous investigators have shown that the faster the rate of presentation of a serial list (Hovland, 1938b), or of a paired-associate list (Hovland, 1949), the greater the facilitation by DP. This relationship (as well as others) has always been a compelling one for the introduction of an inhibitory construct in explanatory attempts. With such a construct it can be said that the faster the rate of presentation the greater the amount of inhibition built up, hence, the greater the positive effect of allowing the inhibition to dissipate over the DP intervals. The question for the moment, however, is how can the rate-of-presentation effect be handled by the present conception?

The slower the rate of presentation the greater the amount of response learning which occurs on any one trial. The greater the amount of response learning the less likely it is that recovered error tendencies will interfere, hence, the less the "need" for extinction. Presenting (at a relatively slow rate) a list in which considerable response integration is required is comparable to presenting a list of fairly well integrated responses at a fast rate. However, the present conception also leads to a prediction in which a slower rate of presentation would give greater facilitation by DP than would a faster rate. If a list of responses is used in which there is very heavy interference in response integration, the fast rate of presentation

will allow so little learning to occur that with certain DP intervals the amount of forgetting would be greater than the positive effects of successive extinctions. A slower rate of presentation—allowing for greater response learning per trial—would give greater facilitation for this DP interval. So far as it is known, there are no data available for this test. To make the test, the following requirements must be met. First, a paired-associate list in which the responses are poorly integrated must be employed. As a guess, a list of poorly integrated trigrams (as presented earlier) would probably be satisfactory, perhaps with the inclusion of a small amount of formal similarity (duplicated letters). Second, the reduction in the rate of presentation must occur in the response inspection period, not in the anticipation period. Thus if the "slow" rate of presentation is 2 seconds for the anticipation period and 2 seconds for the stimulus and response together, a fast rate would be 2 seconds and 1 second, respectively. A reduction in the anticipation interval introduces differences in performance factors which are not included in the present conception. Finally, because of the sharp interaction presumed to occur between amount of interference and the DP interval, several DP intervals should be included. If the above suggested list is used, intervals of 15, 30, and 120 seconds should give full opportunity for the expected reversal in the rate-of-presentation relationship to occur. That is, if it does not occur at 30 seconds, it should occur at 120 seconds.

Length of List

Hovland (1940) has also shown that as the length of a serial list of nonsense syllables increases, a given DP interval facilitates acquisition more and more. One of the inevitable consequences of increasing the length of lists made up of such materials is to increase the amount of formal similarity, and with an increase in amount of formal similarity the greater the likelihood that DP will facilitate acquistion (e.g., Underwood & Richardson, 1958; Underwood & Schulz, 1959). If the amount of interference in response integration was below the optimal amount for the given DP interval, increasing the length of the list would increase the amount of interference toward the optimal. It is apparent that this conception immediately leads to a prediction in which length-of-list relationship as found by Hovland would break down. For, if the shortest list used had the amount of response interference that was optimal for a given DP interval, increasing the length of the list (thus increasing the amount of interference) would lead to less facilitation by DP than was true for the shorter list. No data are available to test this expectation.

Reminiscence in Recall

Reminiscence is demonstrated when recall following a short interval of rest is greater than if no rest occurred. As is well known, this is one of the most elusive phenomena extant in the whole area of verbal learning. Many investigators have been unable to produce the effect. Its magnitude, when found, is very small. For example, in Hovland's (1938a) first study the reminiscence at various points in learning varied from one-third to one-half an item out of a 12-item list. The appearance of the phenomenon is also acutely dependent upon the length of the rest interval (Ward, 1937). Just as the facilitation by DP is dependent upon a highly specific set of conditions, so too is the phenomenon of reminiscence. The present paper has at-

tempted to integrate facts relating DP and learning through the use of certain notions about retention over short intervals. It is reasonable to expect, therefore, that these notions may have some relevance to reminiscence in recall.

When reminiscence is found it is under conditions of rather heavy interlist interference. With serial or paired-associate lists, there is no reason to believe that potential proactive interference tendencies, established by the learning of many earlier lists, are specific to only one item or one association in the list for which reminiscence is demonstrated. That is to say, a proactive error tendency may interfere with more than one item in a list. If this error tendency recovers over a reminiscence interval, it may reach a strength that will interfere with a response but not prevent its correct anticipation. However, with such an error tendency having recovered a certain amount, but with the correct response being anticipated correctly in spite of the recovery, some extinction of that error tendency should occur. If, as presumed above, this error tendency could interfere with other items in the list, especially ones near the anticipatory threshold, a consequence of this partial extinction is to reduce the interference on these near-threshold items. This should give them a slightly higher probability of being correct, thus producing reminiscence.

Again it is clear that this conception of how reminiscence may ocur necessitates a highly specific set of circumstances. The amount of recovery of the error tendency must be "just right," hence the amount of interference and length of interval again become the critical manipulable variables. Furthermore, the interfering tendency must potentially interfere with more than one item so that its partial extinction (when it occurs in conjunction with one response) will reduce the likelihood that it will interfere with other responses. Finally, a single extinction of the interfering tendency must reduce its actual interference effects for other items below those effects which prevail for the same error tendency under the no-rest (control) condition. To suggest that this extraordinarily specific set of conditions must be met before reminiscence can be demonstrated does not seem entirely unreasonable in view of the ephemeral nature of the phenomenon.

Two other comments should be made about reminiscence in recall. First, the above notions have been applied to the situation in which there is interlist interference. In studies of DP, both interlist and intralist interference can produce a positive effect for DP. There appears to be no reason why reminiscence could not be demonstrated with intralist interference with naive subjects, as long as the above mentioned set of conditions are met. That is, there must be interfering tendencies which are not specific to a given response unit, and the amount of interference and length of interval must be coordinated. The present belief is that once these more critical relationships are plotted for DP, it will be a much easier task to specify the exact conditions necessary for the production of reminiscence than is possible at the present time.

The second point to be made is that it is still barely possible that reminiscence is an artifact of the color-naming activity used to fill the interval in many studies. The rationale for this has been outlined elsewhere (Archer, 1953), and in spite of the negative results of Archer's test of this notion, it is conceivable that the conditons necessary to produce this artificial reminiscence are also specific

and that these conditions were not met in the test made.

Reminiscence in Relearning

Certain studies have shown that whereas a single rest interval introduced during the acquisition of a list may not produce enhanced recall after the interval, subsequent learning will be facilitated. Studies of this type are illustrated by the work of Riley (1953, 1954, 1957). In Riley's studies, subjects learned two lists of paired-associate nonsense syllables of low meaningfulness. Since intralist formal similarity was low, interlist similarity among responses was appreciable. A single 2-minute rest interval introduced during the learning of the second list produced some facilitation in subsequent learning although no reminiscence was apparent on the recall trial. No effect was noted when only a single list was learned; therefore, it seems evident that the presence of interlist interference was a critical component for the phenomenon. Furthermore, since the effect occurred whether the two lists had the same stimuli and different responses or different stimuli and different responses, it may also be concluded that the critical interference is not in the stimulus-response relationships. In line with previous work on DP, therefore, it seems likely that the critical locus of interference is in response integration. If this is the case, the single-trial extinction hypothesis can be applied; the interval allows recovery of errors which are extinguished a certain amount; hence, interference with subsequent learning is less than if the interval had not been introduced.

Why does not reminiscence occur on the recall trial immediately after the interval? The reason for this may be that the interference is specific between a response term in the first list and one in the second, e.g., those that begin with the same first letter. If this is the case, extinction of this specific interfering tendency cannot influence the recall of other items on the recall trial since they are interfered with by other specific tendencies.

If this interpretation of Riley's work is appropriate, it would obviously be predicted that no reminiscence effect would be found if the response terms were common one-syllable words. This prediction would hold regardless of the materials used for stimuli and regardless of the relationships between the stimuli in the two lists.

Meaningfulness

There is some evidence that the lower the meaningfulness of the materials the greater the likelihood that DP will facilitate learning. Tsao (1948) concluded this as a result of his experiment, but it is possible that meaningfulness and formal intralist similarity were confounded in his study (Underwood & Richardson, 1958). Nevertheless, in terms of the conception advanced here, meaningfulness should be a variable determining DP effects because in general the lower the meaningfulness the lower the degree of response integration (Underwood & Schulz, 1960). And, as stated several times earlier, the lower the degree of response integration the greater the susceptibility of these responses to interference effects. However, if experiments are performed in which the complete range of meaningfulness is used, but with a single DP interval, the results are not likely to meet expectations. This failure would result from the interaction between amount of interference and length of DP interval. Finally, it should be mentioned again that when interference in integrating responses is very heavy—as it may be with ma-

terial of very low meaningfulness—it is possible that amount of forgetting occurring over even a short DP interval will be too great to expect overall facilitation.

SUMMARY

This paper presents a conception of how distributed practice facilitates the acquisiton of verbal lists. The evidence indicates that distributed practice will enhance learning only when interference occurs in the response-learning phase. This interference reduces the effective response strength of the correct response. With the introduction of a rest interval, error tendencies recover in strength, but with subsequent occurrences of the correct response the error tendency is extinguished. Thus, distributed practice allows for successive extinctions of error tendencies and the assumption is that such a process results in a more effective elimination of the deleterious effects of interference than occurs under massed practice. Under massed practice error tendencies are assumed to be suppressed rather than extinguished.

This conception makes amount of response interference and length of the distribution interval the two critical variables in determining whether or not distributed practice will facilitate learning. In general, the greater the interference the shorter must the distribution interval be for facilitation to occur. If interference is high, and the interval too long, forgetting will occur because of the weak development of the associative strength of the components of the correct response on any one trial and the recovery of the error tendencies will persistently block or replace the correct response. Therefore, for distributed practice to facilitate learning when response interference is heavy, the distribution interval must be short. It is possible that when interference is very heavy, distributed practice will never facilitate learning.

The effects of several other variables, including length of list, rate of presentation, single short rest intervals (reminiscence), and meaningfulness were evaluated. Certain predictions were advanced concerning the effect of these variables when more thoroughly explored.

It should be noted that there are certain "soft" spots in the conception as presented.

1. At the present time the minimal amount of initial interference necessary before a distribution interval of any length will facilitate acquisition cannot be independently specified. The same is true with maximal amount of initial interference.

2. It has been noted that the necessary interference must occur in the response-learning stage; interference in associating the stimulus term with the response term is irrelevant. There is no apparent logical or theoretical reason why the effect of these two sources of interference should be different; however, the data overwhelmingly support the generalization that the interference must be localized in the response term.

3. While the ideas that error tendencies recover with rest, and that they can be extinguished, have independent empirical backing, the notion that recovered error tendencies are more effectively extinguished than nonrecovered tendencies has been developed merely to fit the demands of the data.

The dissatisfaction with these matters may disappear in the research of the next 10 years.

REFERENCES

ARCHER, E. J. Retention of serial nonsense syllables as a function of rest-interval

responding rate and meaningfulness. *J. exp. Psychol.*, 1953, **45**, 245–252.

BARNES, JEAN M. "Fate" revisited. Unpublished PhD dissertation, Northwestern University, 1960.

BARNES, JEAN M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

HOVLAND, C. I. Experimental studies in rote-learning theory: I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224. (a)

HOVLAND, C. I. Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190. (b)

HOVLAND, C. I. Experimental studies in rote-learning theory: VII. Distribution of practice with varying lengths of list. *J. exp. Psychol.*, 1940, **27**, 271–284.

HOVLAND, C. I. Experimental studies in rote-learning theory: VIII. Distributed practice of paired associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.

RILEY, D. A. Reminiscence effects in paired-associate learning. *J. exp. Psychol.*, 1953, **45**, 232–238.

RILEY, D. A. Further studies of reminiscence effects with variations in stimulus-response relationships. *J. exp. Psychol.*, 1954, **48**, 101–105.

RILEY, D. A. The influence of amount of prerest learning on reminiscence effects in paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 8–14.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words, from the Kent-Rosanoff Word Association Test: Studies on the role of language in behavior. Technical Report No. 11, 1954, University of Minnesota, Contract Nonr-66216.

TSAO, J. C. Studies in spaced and massed learning: II. Meaningfulness of material and distribution of practice. *Quart. J. exp. Psychol.*, 1948, **1**, 79–84.

UNDERWOOD, B. J. Studies of distributed practice: III. The influence of stage of practice in serial learning. *J. exp. Psychol.*, 1951, **42**, 291–295.

UNDERWOOD, B. J. Studies of distributed practice: VI. The influence of rest-interval activity in serial learning. *J. exp. Psychol.*, 1952, **43**, 329–340. (a)

UNDERWOOD, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87. (b)

UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149. (a)

UNDERWOOD, B. J. Studies of distributed practice: X. The influence of intralist similarity on learning and retention of serial adjective lists. *J. exp. Psychol.*, 1953, **45**, 253–259. (b)

UNDERWOOD, B. J. Studies of distributed practice: XI. An attempt to resolve conflicting facts on retention of serial nonsense lists. *J. exp. Psychol.*, 1953, **45**, 355–359. (c)

UNDERWOOD, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, **61**, 160–166.

UNDERWOOD, B. J. Studies of distributed practice: XV. Verbal concept learning as a function of intralist interference. *J. exp. Psychol.*, 1957, **54**, 33–40. (a)

UNDERWOOD, B. J. Studies of distributed practice: XVI. Some evidence on the nature of the inhibition involved in massed learning of verbal materials. *J. exp. Psychol.*, 1957, **54**, 139–143. (b)

UNDERWOOD, B. J., & ARCHER, E. J. Studies of distributed practice: XIV. Intralist similarity and presentation rate in verbal-discrimination learning of consonant syllables. *J. exp. Psychol.*, 1955, **50**, 120–124.

UNDERWOOD, B. J., & GOAD, D. Studies of distributed practice: I. The influence of intralist similarity in serial learning. *J. exp. Psychol.*, 1951, **42**, 125–134.

UNDERWOOD, B. J., & RICHARDSON, J. Studies of distributed practice: XIII. Interlist interference and the retention of serial nonsense lists. *J. exp. Psychol.*, 1955, **50**, 39–46.

UNDERWOOD, B. J., & RICHARDSON, J. Studies of distributed practice: XVIII. The influence of meaningfulness and intralist similarity of serial nonsense lists. *J. exp. Psychol.*, 1958, **56**, 213–219.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. Studies of distributed practice: XIX. The influence of intralist similarity with lists

of low meaningfulness. *J. exp. Psychol.*, 1959, **58**, 106–110.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

UNDERWOOD, B. J., & SCHULZ, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. *J. exp. Psychol.*, 1961, **61**, 228–235. (a)

UNDERWOOD, B. J., & SCHULZ, R. W. Studies of distributed practice: XXI. The effect of interference from language habits. *J. exp. Psychol.*, 1961, in press. (b)

WARD, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**(4, Whole No. 220).

WILSON, J. T. The formation and retention of remote associations in rote learning. Unpublished PhD dissertation, Stanford University, 1948.

WRIGHT, S. T., & TAYLOR, D. W. Distributed practice in verbal learning and the maturation hypothesis. *J. exp. Psychol.*, 1949, **39**, 527–531.

(Received September 9, 1960)

Psychological Review
1961, Vol. 68, No. 4, 248–274

VERBAL MEDIATING RESPONSES AND CONCEPT FORMATION[1]

ALBERT E. GOSS[2]

University of Massachusetts

Over the past five decades, verbal mediating responses and stimuli have figured as important elements in a number of stimulus-response analyses of concept formation. This paper briefly reviews these analyses as a prelude to carrying out its main purpose, which is the further explication of the role of verbal mediating responses in conceptual behavior. More specifically, spelled out first are criteria for concept formation tasks, particularly as compared with those for conventional paired-associates tasks. Then described in considerable detail are some paradigms of presumed stimulus-response relationships in concept formation. Finally, the paradigms are considered in conjunction with certain variables and learning principles, and sample predictions are generated.

Probably the earliest explicit stimulus-response analysis of the role of verbal mediating responses in conceptual behavior is that which Max Meyer illustrated with the concept "food" in his *Fundamental Laws of Human Behavior* (1911, pp. 213–214). The same example of the essential features of concepts was used subsequently by Weiss (1925), Dashiell (1928), and Gray (1931). Although Watson (1920, p. 102) chose a different example, his treatment of conceptual behavior also emphasized verbal mediating responses. The primary purpose of these early analyses was to show that conceptual phenomena—which had previously been thought to be impervious to behavioristic treatment—could be dealt with in stimulus-response terms. Understandably for the time, such analyses were only incidentally combined with learning principles to derive predictions about the effects of potentially significant variables on conceptual behavior, and none of the predictions was tested experimentally.

Early in the forties, Birge (1941), Miller and Dollard (1941), and Cofer and Foley (1942) made suggestions concerning the possible significance of verbal mediating responses for conceptual behavior. These treatments, however, were more concerned with defining and applying the mechanism of response-mediated similarity and generalization than with analyzing in detail the role of this mechanism in concept formation. The same is true of Gibson's (1940) development and use of the somewhat parallel notion of internal generalization and its complement, internal differentiation.

A decade later, Osgood (1953) offered *post factum* analyses of the

[1] An earlier version of this paper was presented as part of a symposium on Mediating Processes in Transfer at the 1958 meetings of the American Psychological Association in Washington, D. C.

[2] The proposals in this paper owe much to the ideas and experimental work of Janice E. Carey, James D. Fenn, Harvey Lacey, Marie C. Moylan, and Alvin J. Simmons. An opportunity to read an analysis by Arnold H. Buss aided development of the present treatment of reversal and nonreversal shifts. In their readings of earlier drafts, Barbara S. Musgrave and Charles N. Cofer offered many useful suggestions. Finally, Mary E. W. Goss, though not responsible for the infelicities of presentation which remain, was responsible for the elimination of many others.

conceptual tasks and results described by Hull (1920), Smoke (1932), Heidbreder (1946a, 1946b), and Reed (1946). These analyses, along with those of Baum (1951) and Mandler (1954), emphasized the mechanism of response-mediated similarity and generalization to the virtual exclusion of the complementary mechanism of response-mediated dissimilarity and discrimination. During the same period, Goss and his students extended their studies of the latter mechanism (e.g., Goss & Greenfeld, 1958) to the analysis and investigation of the effects of experimentally controlled verbal pretraining on conceptual sorting (e.g., Fenn & Goss, 1957), conceptual naming (Lacey, 1959), and animistic thinking (e.g., Simmons & Goss, 1957).

Hypotheses about the role of verbal mediating processes in reversal and nonreversal shifts of conceptual phenomena have been proposed and tested by Kendler and his students (e.g., Kelleher, 1956; Kendler & D'Amato, 1955), as well as by Buss (1956), Gormezano and Grant (1958), and Harrow and Friedman (1958). These proposals apparently evolved primarily from the considerable body of data and theory concerning the simple discriminative behaviors of infrahuman and preverbal human organisms to which Spence (1936) has been the major contributor, rather than from existing hypotheses and data concerning response-mediated similarity and dissimilarity. Of similar origin is Wicken's (1954) analysis of the strengthening of discriminative responses to values along one dimension of multidimensional stimuli, and his subsequent more explicit hypotheses as to how verbal mediating responses might be the vehicles of "perceptual sets" (Wickens & Eckstrand, 1954).

Pavlov's "second signal system" is essentially equivalent to mediating responses and stimuli, and it has been the basis for recent analyses of "higher nervous activity" by Soviet psychologists (e.g., Elkonin, 1957). Within this framework, Liublinskaya (1957) has described theoretical and experimental work on the role of the second signal system in the conceptual behaviors of preschool children.

Of the many analyses and studies that bear directly or indirectly on the role of verbal mediating responses and stimuli in concept formation, some have been supported solely by informal examples rather than by experimental data and principles. Those which do refer to experimental materials have often been limited to one or two relatively specific situations. And there has been a tendency to consider the nature and implications of only a few of the many possible patterns of relationships that can exist among initiating stimuli, mediating responses and stimuli, and terminating responses (Goss, 1955, 1956).[3] There is clearly need for a

[3] A temporal sequence of stimulus-response events in which a mediating response and stimulus may be distinguished can be represented as $S_{\text{Initiating}}-R_{\text{Mediating}} \sim S_{\text{Mediating}}-R_{\text{Terminating}}$. Social situations or experimental tasks are conceived as beginning with or initiated by some stimulus "element" or "compound" and as terminating with a response which is reinforced or punished; is instrumental in altering a subject's environment; or, more generally, has simply been designated the terminating, reference, or criterion response. Any stimulus event or receptor activation might be the initiating stimulus of a sequence, though usually and practically such stimuli are social and physical events.

Ideally, two criteria must be met in order for responses and the stimuli they produce to be considered mediating responses and stimuli. The first criterion is the observation of or grounds for inferring the occurrence of one or more responses subsequent to the initiating stimulus and before the terminating response. The second criterion is the demonstration that such temporally intermediate responses

more comprehensive yet experimentally rooted analysis; within the limitations to be described, such an analysis is offered in this paper.

Criteria for Concept Formation Tasks

General specification of the nature of concept formation tasks is a logical starting point. Because many concept formation tasks have much in common with conventional paired-associates learning tasks, differentiation of the two types of tasks is also required.

General Criteria

Fundamental to the definition of concept formation tasks (conceptual behaviors) are patterns of relationships between initiating stimuli and terminating responses. More particularly, such tasks involve patterns in which two or more independently presented initiating stimuli evoke the same terminating response. It is the independent presentation of stimuli that distinguishes concept formation tasks from convergent stimulus-compound situations. Thus crudely characterized, of course, the simplest concept formation tasks are essentially identical to phenomena more often labeled primary stimulus generalization and response-mediated generalization (Dollard & Miller, 1950; Goss, 1955). Indeed, the latter phenomena might be looked on as limiting cases of the former.

Those situations *commonly* regarded as concept formation tasks, however, are more complex. Sets of initiating stimuli are partitioned into two or more subsets, at least one of which has two or more independently presented members. Usually each of the subsets has two or more members, and the learning requirement is acquisition of the same response to all members of a particular subset and of a different response for each subset.

At a descriptive level, the sets of initiating stimuli in concept formation studies have been markedly heterogeneous. Because some sets apparently require paradigms different from those for other sets, and also for simplicity, sets of initiating stimuli are divided here into three types. These three types seem sufficient both for the development of adequate one-stage and two-stage paradigms and for the representation of all sets of initiating stimuli.

In the first type of set, all members are either variations in values along one physical or psychophysical dimension, or they are combinations of values along two or more dimensions. The dimensions may be primary or derived; the combinations may be completely or incompletely orthogonal. Illustrative of such sets are four squares which are red-small, red-large, blue-small, and blue-large.

Initiating stimuli in the second type of set can be partitioned into two or more subsets on the basis that all stimuli of each subset have some physically specifiable element or relation in common. The stimuli within each of these subsets differ from each other with respect to additional fea-

and stimuli have actual or potential facilitative or inhibitory effects on one or more measures of the occurrence and strength of the terminating response. Relative covertness and some particular topography as additional criteria seem unnecessarily restrictive. However, because of the presumed greater functional significance for most complex behaviors of postverbal humans of mediating responses originally or presently involving the vocal musculature, the focus of the analyses developed here will be on such responses and the stimuli they produce—verbal mediating responses and stimuli. A more exhaustive treatment of the definition of mediating responses and stimuli and of the bases for inferring or confirming their occurrences and effects can be found in Goss (1956).

tures. Thus, the stimuli of each subset consist of both common and variable features, neither of which has been (or perhaps could be) completely reduced to combinations of physical or psychophysical dimensions. Four stimuli, two of which have an S-shaped form in common and two of which have a sword-shaped form in common, but whose other features differ, are representative of this type of set of initiating stimuli.

Sets of initiating stimuli which are less readily or not at all reducible to combinations of values along dimensions, or to subsets defined by common elements or relations, constitute the third type of set. Illustrative of this type are sets of words for objects, properties, or relations. Subsets of words are usually, but not necessarily, specified on the basis of observations or assumptions that all of the stimuli of each subset evoke one or more common responses, some of which differ from the common responses evoked by the stimuli of each of the other subsets. An example of such sets of initiating stimuli is provided subsequently.[4]

With this type of initiating stimuli the bases for partitioning into subsets and for assigning responses to those subsets might be entirely arbitrary or random. For example, eight consonant-vowel-consonant initiating stimuli, none of which has any letters in common, might be randomly partitioned into four subsets of two members each. As stimuli for responses, a different one of four two-digit numbers, none of which has any digit in common, might then be randomly assigned to each of the subsets of initiating stimuli, with the requirement that a different response be conditioned to each subset of initiating stimuli.

Paired-Associates Learning Tasks and Concept Formation Tasks

Paired-associates learning can be regarded as referring either to a particular kind of task or to a more general *procedure* for establishing and changing stimulus-response associations. Many concept formation *tasks*, however, have employed the paired-associates *procedure* for strengthening associations between stimulus members and responses elicited by response members. Both conventional paired-associates tasks and such concept formation tasks may therefore be regarded as complementary special cases of patterns of stimulus-response associations which are strengthened by the paired-associates procedure (Metzger, 1958; Richardson, 1958).

There is only one essential difference between conventional paired-associates learning tasks and concept formation tasks in which stimulus-response associations are established by the paired-associates procedure. That difference is in the ratio of stimulus members to responses which are to be conditioned to those stimuli. For conventional paired-associates learning tasks, the ratio of stimulus members to response members has been 1:1: i.e., separate associations

[4] The first and possibly the second of the three types of sets of initiating stimuli distinguished here and the relationships with terminating responses into which these types of stimuli enter are equivalent to what have been labeled elsewhere as conjunctive categories or concepts (Bruner, Goodnow, & Austin, 1956, pp. 41–43, 244–245). The third of the present types seems approximately equivalent to Bruner, Goodnow, and Austin's disjunctive categories or concepts. From Bruner, Goodnow, and Austin's definition of relational concepts or categories and the accompanying examples, it cannot be determined whether such relational categories overlap with the first and second of the types noted here or whether such categories involve some additional type of initiating stimuli not distinguished here.

are established between each of mn_s different stimulus members and each of the mn_r different responses elicited by mn_r response members.

For the formation of concepts by the paired-associates procedure, however, the ratio of stimulus members to the responses which are conditioned to those stimuli has been greater than 1:1: i.e., for at least one, and usually for all of m subsets of stimulus members, $n_{s_j} > 1$, where n_{s_j} is the number of stimuli in the jth subset. Regardless of the type of sets of initiating stimuli, by increasing the numbers of responses to equal the number of initiating stimuli, concept formation tasks in which stimulus-response associations are established by the paired-associates procedure can be transformed into conventional paired-associates learning tasks. Conversely, by decreasing the number of responses from equality with the number of initiating stimuli the latter can be transformed into concept formation tasks.

PARADIGMS

The role of verbal mediating responses in concept formation tasks can be developed most easily and clearly by means of two-stage paradigms of presumed relationships among initiating stimuli, mediating responses and stimuli, and terminating responses for each of the three types of initiating stimuli which were distinguished in the preceding section. Inferences regarding mediating responses and stimuli are usually based on characteristics of relationships between initiating stimuli and terminating responses. Accordingly, in the first part of this section, the two-stage paradigms are developed within the framework of one-stage paradigms which involve only initiating stimuli and terminating responses. Noted in connection with the description of these paradigms are some explanatory consequences, in particular, for reversal and nonreversal shifts.

Concept formation tasks are usually complex, and mediating responses and stimuli are commonly inferred rather than observed directly. Two-stage paradigms of conceptual behaviors should, therefore, be proposed cautiously. Emphasized in the second part of this section are some precautions in the development and use of two-stage paradigms.

"Abstract set or attitude," "hypotheses," and "strategies" are notions often advanced as central to any explanations of conceptual behaviors. Moreover, they are often regarded as opposed to stimulus-response analyses of concept formation tasks. The thesis elaborated in the last part of this section, however, is that these are not opposing notions, but rather are already embodied or can be readily assimilated within the one-stage and two-stage paradigms of the present analysis.

One-Stage and Two-Stage Paradigms

One-stage paradigms of conceptual situations and behaviors involve relationships between initiating stimuli and terminating responses. Such paradigms provide baselines for the development of two-stage paradigms, which introduce verbal mediating responses and stimuli. One-stage paradigms are not merely stepping stones, however; they are useful in themselves, in that they appear to represent adequately some of the conceptual behaviors of infrahuman organisms, of preverbal humans, and of humans under conditions which preclude or short circuit verbal mediating responses.

Combinations of Values along Dimensions

One-stage paradigms. A set of stimuli consisting of combinations of two values along each of two dimensions is the simplest case of possible sets of stimuli containing complete orthogonal combinations of m values along each of n dimensions. The four initiating stimuli of the one-stage paradigm shown in Figure 1 are combinations of two values (x_1, x_2) along an X dimension, and of two values (y_1, y_2) along a Y dimension. For example, x_1 and x_2 might be the values giving rise to the colors red and blue, respectively, along a dimension of wave length; y_1 and y_2 might be small and large areas, respectively, along a (derived) dimension of size.

The two terminating responses could be naming by means of familiar words, nonsense syllables, or manipulanda representing two different names. Or they could be sorting by placing the stimuli in groups, matching them with other stimuli, or approaching-avoiding. The two patterns of relationships between initiating stimuli and terminating responses depict the associations whose strengthening or occurrence are referred to here as concept formation. That both animals and humans can acquire such differential responses to one or some of the dimensions of multidimensional stimuli has been amply demonstrated (Kelleher, 1956; Kendler & D'Amato, 1955; Woodworth, 1958).

Determinants of the actual and potential patterns of relationships that will be learned include the number of fixed, relevant, and irrelevant dimensions presented, as well as the number of values selected along the relevant and irrelevant dimensions. A *fixed* dimension is exemplified by but one value along the dimension for all of the members of the set of initiating

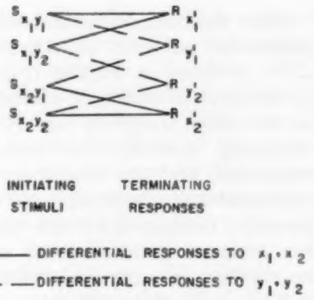

FIG. 1. One-stage paradigm of relationships involving initiating stimuli which are combinations of two values, x_1, x_2 along an X dimension and two values, y_1, y_2 along a Y dimension. (The differential terminating responses $R_{x'_1}$, $R_{x'_2}$ are to x_1, x_2, while $R_{y'_1}$, $R_{y'_2}$ are to y_1, y_2. The X dimension might be color with values of red, x_1, and blue, x_2, and the Y dimension might be size with values of small, y_1, or large, y_2.)

stimuli. For the stimuli shown in Figure 1, form, dimensionality, number of forms, and location of the forms on the presentation cards might be the same for each of the four initiating stimuli. They are among the fixed dimensions of those stimuli.

Should the task be to respond on the basis of red or blue, disregarding size, or to respond on the basis of small and large, disregarding color, the relevant dimensions would be color and size, respectively. More generally, the dimensions of the combinations of values to which discriminative responses are to be conditioned are the *relevant* dimensions. *Irrelevant* dimensions are those which, in the formation of some particular concept(s), involve values which must be disregarded. Such dimensions—or, more precisely, the values along such dimensions—may be completely or incompletely orthogonal with respect to combinations of values along the relevant dimensions.[5]

[5] As is suggested by the overlap of the terminology employed here and that employed

Once the component associations of particular patterns of relationships, such as those in Figure 1, are at a given level of strength, changes may occur either *singly* or *jointly*, in the initiating stimuli, the terminating responses, and the stimulus-response relationships. Such changes are important, because they are the bases for concept generalization and for reversal and nonreversal shifts.

The initiating stimuli can be changed by adding or shifting to new values along the original dimensions of the stimuli, or by adding or subtracting dimensions and values along those dimensions. The degree of occurrence of previously learned differential responses to altered sets of stimuli is the measure of *concept generalization*. Except where concept generalization has been used as a criterion of concept formation (e.g., Heidbreder, 1946a, 1946b), however, this phenomenon has not been of great experimental interest. For this reason, to elaborate on concept generaliztaion here is considered premature.

Both the initiating stimuli and the terminating responses can remain the same, but their relationships, or the relevant and irrelevant dimensions and values, can be changed by reversal or nonreversal shifts. The effects of such shifts on conceptual behaviors, and explanations of those effects, have been among the major concerns of many recent studies of concept formation (e.g., Kendler & Kendler, 1959). It is important, therefore, to describe reversal and nonreversal shifts within one-stage paradigms for this type of initiating stimuli. Also, such description is prerequisite to the subsequent analysis of the role of verbal mediating responses and stimuli in reversal and nonreversal shifts.

With reversal shifts the values or combinations of values to which differential responses are learned remain the same, but the responses to values or combinations of values are interchanged. In Figure 1, for example, $R_{x'_1}$ might be shifted from $S_{x_1y_1}$ and $S_{x_1y_2}$ to $S_{x_2y_1}$ and $S_{x_2y_2}$; and $R_{x'_2}$ would become the reinforced response to $S_{x_1y_1}$ and $S_{x_1y_2}$ instead of to $S_{x_2y_1}$ and $S_{x_2y_2}$. Specifically, the response to red-small and red-large would be shifted to blue-small and blue-large, and the response to blue-small and blue-large would be made to red-small and red-large.

A complete nonreversal shift entails a change from differential pairings of responses with combinations of values along one or more dimensions to differential pairings of those responses with combinations of values along one or more entirely different dimensions. Thus, the pattern of relationships in Figure 1 might be changed from responding in terms of x_1 and x_2 along X, disregarding y_1 and y_2 along Y, to responding differentially to y_1 and y_2, disregarding x_1 and x_2. Only the relationships of the two responses to the initiating stimuli and not the responses themselves are changed. The relationship between $S_{x_1y_1}$ (red-small) and $R_{x'_1}$ would remain the same but that response would be changed from $S_{x_1y_2}$ (red-large) to $S_{x_2y_1}$ (blue-small). The relationship between $S_{x_2y_2}$ (blue-large) and $R_{x'_2}$ would remain the same but that response would be changed from $S_{x_2y_1}$ (blue-small) to $S_{x_1y_2}$ (red-large).

New terminating responses can be introduced. Should the old and the new responses have the same topography and, because of time limitations imposed by the task, be prohibited from occurring in sequence, the old

in classifying analysis of variance designs (e.g., Federer, 1955), such designs provide models of some of the many possible relationships between terminating responses and sets of initiating stimuli which are combinations of values along dimensions.

responses must be inhibited for the new responses to occur. Such a state of affairs has been described as a condition, if not the optimum condition, for *negative transfer*. What results is simply a shift from one one-stage paradigm to another one-stage paradigm. But if the old and new responses do not interfere with each other (have separate topographies or can occur in sequence), the old responses may not drop out but instead constitute relatively stable links—mediating responses and stimuli—between initiating stimuli and the new terminating responses. Thus, a two-stage paradigm would have emerged. This is, of course, the sequence of events which has been presumed in investigations of the effect of verbal pretraining on subsequent conceptual sorting and naming (e.g., Fenn & Goss, 1957).

Despite the usefulness and greater simplicity of one-stage paradigms, there are considerations which suggest that such paradigms are less adequate than two-stage paradigms for explanation and prediction of the conceptual behavior of verbal humans in many concept formation tasks and even, perhaps, of some of the conceptual behaviors of infrahuman organisms and nonverbal humans. These considerations include: (*a*) observations of positive transfer from verbal pretraining to subsequent conceptual sorting or naming and of facilitation due to instructions or instruction induced sets (Carey & Goss, 1957; Fenn & Goss, 1957; Gelfand, 1958; Goss & Moylan, 1958; Hunter & Ranken, 1956; Lacey & Goss, 1959), (*b*) the relatively greater ease of reversal than of nonreversal shifts for human adults (Buss, 1956; Gormazano & Grant, 1958, Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Mayzner, 1956) and for children who are fast learners (Kendler & Kendler, 1959) in contrast to the superiority of nonreversal shifts for animals (Kelleher, 1956) and for children who are slow learners (Kendler & Kendler, 1959), and (*c*) verbal humans' reports of the occurrence and use of names for dimensions and values of stimuli in the conceptual sorting of stimuli (e.g., Lacey & Goss, 1959). An additional consideration rests primarily on the results of studies employing the third type of sets of stimuli (e.g., Reed, 1946). Without the postulation of common verbal or other responses to subsets of stimuli whose members are highly dissimilar physically, generalization of a common terminating response from one stimulus of a subset to other stimuli of the subset would be precluded. Each of the associations between initiating stimuli and terminating responses would have to be strengthened separately, with a consequent increase in difficulty of mastering the task.

Two-stage paradigms. Shown in Figure 2 are some of the possible stimulus-response relationships in a two-stage expansion of the one-stage paradigm presented in Figure 1. The four subsets of these relationships which should be distinguished are those: between initiating stimuli and mediating responses; between mediating stimuli and mediating responses or, more simply, between mediating responses; between mediating stimuli and terminating responses; and between initiating stimuli and terminating responses.

Within the first of these subsets of relationships, variations in the strength of two subpatterns of relationships between initiating stimuli and mediating responses may have somewhat different effects on conceptual behaviors. The first subpattern represents relationships in which responses of naming the dimensions

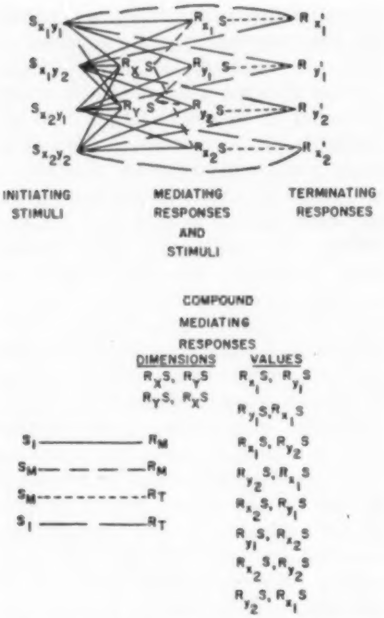

FIG. 2. Two-stage paradigm of some of the relationships possible between initiating stimuli and mediating responses, between mediating stimuli and mediating responses, between mediating stimuli and terminating responses, and between initiating stimuli and terminating responses. (In order to simplify the possible relationships of the diagram, relationships involving compound mediating responses for dimensions and compound mediating responses for values along dimensions are listed separately. The mediating responses might be differential with respect to x_1, x_2 along the X dimension or to y_1, y_2 along the Y dimension which is also the case for the terminating responses.)

occur. These are the associations between the initiating stimuli and R_X, R_Y. The second subpattern represents responses of naming the specific values along the dimensions. The responses of these associations are R_{x_1} for x_1; R_{x_2} for x_2; R_{y_1} for y_1; and R_{y_2} for y_2.

When the relationships between mediating stimuli and mediating responses are added, variations in the strengths of three more subpatterns of relationships can be distinguished. The first of these subpatterns is sequences of responses of naming the dimensions. These appear in the lower half of Figure 2 under "Dimensions" as R_XS, R_YS and R_YS, R_XS. The second subpattern is sequences of responses of naming values along dimensions. The eight sequences of combinations and orders of two of such responses are shown in the lower half of Figure 2 under "Values." The third of these subpatterns is sequences of responses of naming both dimensions and values along dimensions. For example, combining one of the two responses of naming a dimension with one of the four responses of naming a value would generate 16 permutations of a particular dimension response with a particular value response.

Variation in the strength of each of these five subpatterns of relationships between initiating stimuli and mediating responses might have somewhat different effects on the direction and degree of: extralist response interference with both mediating responses and terminating responses, trial-to-trial variability of the stimulus patterns immediately prior to the terminating responses, response-mediated similarity, and response-mediated dissimilarity. In turn, these conditions should influence direction and degree of transfer to acquisition of associations between initiating stimuli and terminating responses. Table 1 summarizes assumptions about the effects of each of the first four subpatterns on extralist response interference, trial-to-trial variability of stimulus patterns, response-mediated similarity, and response-mediated dissimilarity. Table 1 also indicates whether these four consequences of each of the four subpatterns considered separately are expected to be

facilitative (+), inhibitory (−), or neutral with respect to the formation of particular concepts. At present there is no way of combining the separate presumed facilitative, inhibitory, or neutral effects into a net facilitative, inhibitory, or neutral effect.

Except where precluded by prior training in the experimental situation, by selection on the basis of associations to the same or similar sets of initiating stimuli, or by instructions, each of these four subpatterns might occur both within trials and in successive trials during a good part of the course of acquiring associations between the initiating stimuli and the terminating responses. Their relative strengths at any point in learning—and, therefore, their effects on acquisition of initiating stimulus-terminating response associations—will be contingent on factors which include the following: their initial relative strengths, the values or combinations of values along one or more dimensions to which the differential terminating responses are being strengthened, time permitted to make the terminating responses, and degree of mastery of the terminating responses.

The fifth subpattern, which involves sequences of mediating responses of naming the dimension and of naming values along the dimension, may also influence acquisition of terminating responses. For example, fairly strong bidirectional associations might exist or be established between R_XS and R_{x_1}, R_{x_2}, and between R_YS and R_{y_1}, R_{y_2}. Should R_X be stronger than R_Y, R_{x_1} and R_{x_2} would occur and be available for mediating discriminative terminating responses to $S_{x_1y_1}$, $S_{x_1y_2}$ and to $S_{x_2y_1}$, $S_{x_2y_2}$ rather than to $S_{x_1y_1}$, $S_{x_2y_1}$, and $S_{x_1y_2}$, $S_{x_2y_2}$. Contingent on the relationships between initiating stimuli and terminating responses which were being differentially reinforced, facilitation or inhibition of these associations might be occasioned.

The remaining two subsets of relationships, those between mediating stimuli and terminating responses and those between initiating stimuli and terminating responses, are of primary

TABLE 1

SPECIFIC PATTERNS OF STIMULUS-RESPONSE RELATIONSHIPS

Specific Patterns Involving	Extralist Response Interference	Stimulus Variability	Response-Mediated Similarity	Response-Mediated Dissimilarity
Dimensions:				
R_X, R_Y	Reduce (+)	Reduce (+)	Increase (−)	Decrease (−)
R_XR_Y, R_YR_X	Reduce (+)	Reduce (+)	Increase (−)	Decrease (−)
Values along single dimensions:				
R_{x_1}, R_{x_2}	Reduce (+)	Reduce (+)	Increase for $S_{x_1y_1}$, $S_{x_1y_2}$ and for $S_{x_2y_1}$, $S_{x_2y_2}$ (+ for $R_{x'_1}$, $R_{x'_2}$ and − for $R_{y'_1}$, $R_{y'_2}$)	Increase for $S_{x_1y_1}$, $S_{x_1y_2}$ in relation to $S_{x_2y_1}$, $S_{x_2y_2}$ (+ for $R_{x'_1}$, $R_{x'_2}$ and − for $R_{y'_1}$, $R_{y'_2}$)
R_{y_1}, R_{y_2}	Reduce (+)	Reduce (+)	Increase for $S_{y_1x_1}$, $S_{y_1x_2}$ and for $S_{y_2x_1}$, $S_{y_2x_2}$ (+ for $R_{y'_1}$, $R_{y'_2}$ and − for $R_{x'_1}$, $R_{x'_2}$)	Increase for $S_{y_1x_1}$, $S_{y_1x_2}$ in relation to $S_{y_2x_1}$, $S_{y_2x_2}$ (+ for $R_{y'_1}$, $R_{y'_2}$ and − for $R_{x'_1}$, $R_{x'_2}$)
Combinations of values along dimensions:				
$R_{x_1}R_{y_1}$, $R_{y_1}R_{x_1}$, etc.	Reduce (+) to Increase (−)	Reduce (+) to Increase (−)	No differential effects among initiating stimuli	

Note.—The direction of these effects are shown along with whether they are expected to have facilitative (+) or inhibitory (−) consequences. In the case of responses to values along single dimensions, whether particular initiating stimulus-mediating response relationships are facilitative or inhibitory is contingent on the relationships between initiating stimuli and terminating responses which are to be acquired.

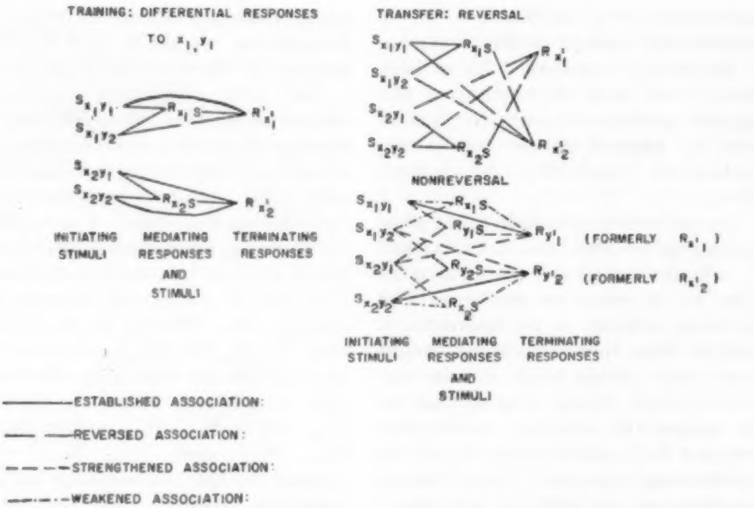

FIG. 3. The paradigm for training shows conceptual responses, $R_{x'_1}$ and $R_{x'_2}$ to $S_{x_1y_1}$, $S_{x_1y_2}$ and $S_{x_2y_1}$, $S_{x_2y_2}$, respectively, after whose strengthening to some criterion level, reversal and nonreversal shifts are made. (As shown in the two paradigms for transfer, a reversal shift requires changes in only 6 associations, while a nonreversal shift may affect up to 14 associations. The terminating responses remain the same with respect to general topography and specific form—only the stimulus-response relationships into which they enter are altered. $R_{y'_1}$ and $R_{y'_2}$ of the nonreversal shift are the same responses as $R_{x'_1}$ and $R_{x'_2}$, respectively. However, the subscripts were changed to indicate that differences along the Y dimension, y_1, y_2, are the new bases for differential responses.)

importance here because of their presumed roles in reversal and nonreversal shifts. The upper diagram of Figure 3 shows the relationships among initiating stimuli, mediating responses and stimuli, and terminating responses which might exist at appreciable levels of strength upon attainment of differential responses to the x_1 and x_2 values along the X dimension. Should there be introduction of differential reinforcement of R_{x_2} to $S_{x_1y_1}$, $S_{x_1y_2}$, and of R_{x_1} to $S_{x_2y_1}$, $S_{x_2y_2}$ to bring about a reversal shift, six associations might be changed: the four between the initiating stimuli and the terminating responses, and the two between the mediating stimuli and the terminating responses. In contrast, 14 associations might be affected by a nonreversal shift to the differential reinforcement of responses to the y_1 and y_2 values of Y. These are the four between initiating stimuli and R_{x_1}, R_{x_2}, which might be weakened; the four between those stimuli and R_{y_1}, R_{y_2}, which might be strengthened; the two between S_{x_1} and S_{x_2} and the terminating responses which might be extinguished while the two between S_{y_1} and S_{y_2} and those responses are established and strengthened; and the two associations between initiating stimuli and terminating responses ($S_{x_1y_2}$ and $R_{x'_1}$; $S_{x_2y_1}$ and $R_{x'_2}$), which must be reversed.

If equal weights are assumed for the component associations of two-stage paradigms, and if shifting is inversely related to the number of associations which must or may have to be changed, reversal shifts should be accomplished

more rapidly than nonreversal shifts. Within one-stage paradigms, reversal shifts will affect more associations and therefore be more difficult than nonreversal shifts. Thus, as Kendler and his associates (Kendler & D'Amato, 1955) have suggested, but without detailed development of the basis for this proposal, one-stage and two-stage paradigms generate opposing predictions about the relative ease of reversal and nonreversal shifts.

In general, as the number of dimensions and values increases, the number of associations involved in nonreversal shifts becomes increasingly greater than the number altered by reversal shifts. Other factors equal, therefore, with greater numbers of dimensions and values, the relative disadvantage of nonreversal shifts should become greater. Other factors, however, are not likely to be equal. As Buss (1956) has noted, with nonreversal shifts responses on the basis of the previously reinforced values along the no-longer-relevant dimension continue to be reinforced on 50% of the trials. The weakening of such differential responses will, therefore, be retarded, and will further contribute to the disadvantage of nonreversal shifts. But four additional conditions may serve to reduce the relative disadvantage of nonreversal shifts.

First, once some of the initiating stimuli begin to elicit the new mediating responses, when other initiating stimuli are presented these mediating responses should generalize extensively among those stimuli. One basis of such generalization would be the presence of stimuli common to each trial: i.e., those arising from the experimental situation, from postural and receptor-orienting responses, and from responses to instructions other than those aspects referring to more specific associations between initiating stimuli and mediating responses.

Second, because each of the values along the new dimension of the nonreversal shift is an element common to a subset of stimuli, the new mediating response for a subset of initiating stimuli should generalize among the members of the subset. Simultaneously, of course, the same two conditions should result in the generalization of inhibition of the relationships between initiating stimuli and the old mediating responses.

Third, though not included in the paradigms of Figure 3, the response R_X, which represents the response of naming the X dimension, might be replaced by the comparable response, R_Y, for the Y dimension. The increased frequencies of occurrence of stimuli produced by R_{y_1} and R_{y_2}, which are presumably already associated with R_Y, would be the bases of the initial evocations of R_Y. Because R_Y is only one response, however, its strengthening and generalization among the initiating stimuli should be even faster and more extensive than the strengthening and generalization of R_{y_1} and R_{y_2}. Therefore, R_Y should begin to occur first and, because of the pre-established associations between the stimuli it produced and R_{y_1} and R_{y_2}, their probabilities of occurrence relative to R_{x_1} and R_{x_2} should be increased markedly.

Fourth, successive reversal or nonreversal shifts should increase the probabilities of occurrence of mediating responses referring to a change in task. With successive reversal and nonreversal shifts, number of trials to learn the new concepts of each shift usually decreases. Some of this increasingly more rapid formation of new concepts is probably due in part to warm up or performance set in the form of familiarization with mode and rate of presentation of the initiating stimuli. Such familiarization should

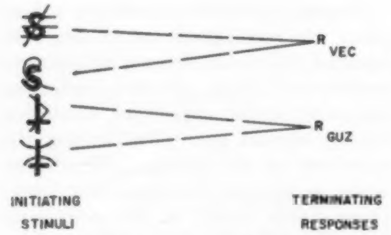

FIG. 4. One-stage paradigm of relationships between nonsense syllable terminating responses and initiating stimuli which consist of two different common elements each of which is accompanied by features which differ from figure to figure.

eliminate irrelevant competing responses as well as lead to greater stability of postural and receptor-orienting responses to thus assure more effective reception of the initiating stimuli and lower the variability of response-produced stimuli. Also, with experience, subjects should learn to recognize with greater confidence and greater accuracy that they have reached perfect or near perfect performance of discriminative responses to some subsets of stimuli. Consequently, any error then made would serve as a cue that the experimenter has shifted the concepts rather than the concepts have not yet been learned. Further, subjects will be increasingly familiar with whether the shifts are reversal or nonreversal shifts and, if the latter, with how many dimensions have probably been shifted and even to what dimensions the shifts have probably been made. Thus, mediating responses of the form "He's changed the task" or "Something has changed" should come to control whole sets of further mediating responses which name dimensions and values along dimensions. These four conditions should reduce the net disadvantage of nonreversal to reversal shifts to margins which are much less than those suggested by simply counting the numbers of equally weighted associations which such shifts might affect.

Common Elements or Relations

One-stage paradigms. Figure 4 is a one-stage paradigm for concept formation with sets of stimuli, such as those constructed by Hull (1920), in which each subset requiring a common response consists of a common element accompanied by other features which vary unsystematically from instance to instance. Furthermore, the common elements are neither completely nor incompletely orthogonal combinations of values along one or more discernible physical or psychophysical dimensions. If common "relations" among the parts of complex forms are regarded as separable from the features which vary among instances exemplifying the same relation, Smoke's (1932) set of initiating stimuli and other sets that resemble his can also be represented by this paradigm.

Two-stage paradigms. The two-stage paradigm for the second type of initiating stimuli is shown in Figure 5. The letters or word subscripts of the mediating responses are

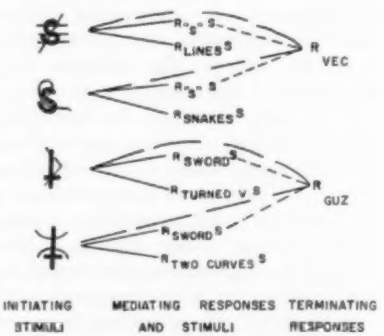

Fig. 5. Two-stage paradigm of the relationships possible between initiating stimuli, which consist of two different common elements and variable features, and both mediating and terminating responses and also between mediating stimuli and terminating responses.

possible specific pre-established mediating responses to the indicated initiating stimuli. Worth noting, because of their predictive consequences, are three major differences between this paradigm and the two-stage paradigm for sets of stimuli composed of combinations of values along dimensions shown in Figure 2.

First, no responses of naming the component dimensions are present. However, should the common elements of two or more subsets be at the same spatial position, responses of orienting-toward and naming that position might occur and be strengthened. Although such responses and the stimuli they produce would be nondifferential with respect to the relationships between initiating stimuli and terminating responses, their occurrence might reduce extralist intrusions and stimulus variability as well as assure more frequent reception of the elements which distinguish one subset of figures from another.

Second, both the common element of members of a subset and the variable features of those members are likely to be made up of a fairly large number of discriminable features, each of which elicits naming responses. If the common element or relation which defines a particular subset of stimuli does elicit some response which is the same for all members of the subset, that response to each member is likely to have considerable competition from the responses to other parts of the common element as well as from responses to the variable features of each member. Such responses and the further responses which they may evoke may interfere, not only with any common mediating response to all stimuli of a subset, but also with the terminating response for that subset. Further, few, if any, pre-established stable patterns of associations among mediating responses might exist. Such conditions should also foster high trial-to-trial stimulus variability.

Third, the variable features of stimuli with a given common element or relation may have little or no physical similarity to those features of the other subsets of stimuli with common elements. And, as already suggested, the naming responses evoked by one subset may have little overlap with those evoked by the other subsets. While reversal shifts could be instituted, such characteristics of the stimuli would severely limit or obviate nonreversal shifts.

Elicitation of Common Responses

One-stage paradigms. The third kind of stimuli are those whose subsets are distinguished from each other on the basis of their members' elicitation of some common response that differs from the common responses defining each of the other subsets. Figure 6 is a one-stage paradigm and Figure 7 is a two-stage paradigm of the stimulus-response relationships presumed to be involved in the formation of concepts with such stimuli.

In the one-stage paradigm, increased frequency of arousal of each one of the common terminating responses by the stimuli of the subset is viewed as strengthening of the concept. The stronger the initial associations between the stimuli of subsets and their terminating response, and the higher the variance of those associations, the more rapid the formation of concepts (Freedman & Mednick, 1958; Underwood & Richardson, 1956). The limiting case of the formation of such concepts is the acquisition, from zero levels of initial strength, of common responses which have each been assigned to a different subset of physically dissimilar stimuli.

Two-stage paradigms. The mediating responses of the two-stage para-

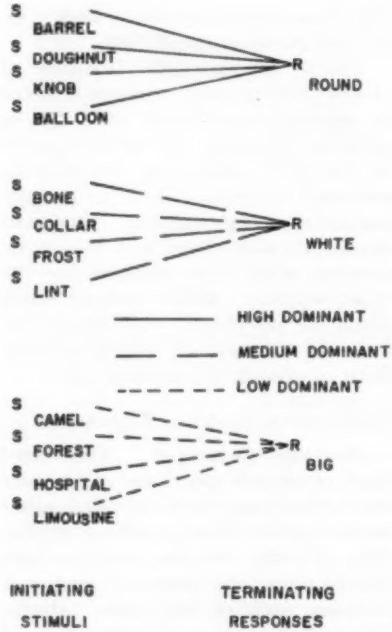

FIG. 6. One-stage paradigm of relationships between subsets of initiating stimuli each of which is defined by elicitation of a common terminating response by the stimuli. (The words used are from Underwood and Richardson's—1956—first list and are of three levels of dominance with respect to their elicitation of the terminating responses.)

digm (Figure 7) are those which define each of the subsets of initiating stimuli. In general, though not necessarily, the associations between initiating stimuli and terminating responses would be at zero levels initially, as would those between mediating stimuli and those responses. To the degree that each stimulus of a subset elicits the common mediating response, there will be a response-mediated increase in the similarity of those largely dissimilar stimuli; and, once the mediating stimulus is associated with the terminating response, acquisition of the concept should be facilitated by response-mediated generalization. Griffith and Spitz (1958) and Griffith, Spitz, and Lipman (1959) obtained direct relationships between correct abstractions made by normal and mentally retarded children and number of words defined by the same possible common abstractions. From these relationships they inferred that the common definition mediated the common abstractions.

The presence of different mediating response-produced stimuli, each associated with a different terminating response, might also increase the response-mediated dissimilarity and discrimination of stimuli which are members of different subsets (Fenn & Goss, 1957). Such an increase would counteract any generalization of terminating responses among sub-

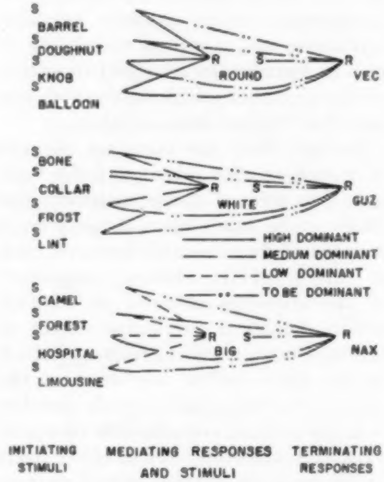

FIG. 7. Two-stage paradigm of relationships possible between subsets of initiating stimuli defined by the elicitation of common responses which now function as mediating responses and stimuli in the acquisition of new associations between each of the subsets of initiating stimuli and their common nonsense syllable terminating responses. (The mediating responses should increase the similarity of stimuli within subsets and decrease the similarity of stimuli of each subset to those of the other subsets.)

sets due to fortuitous physical resemblances among stimuli belonging to different response-defined subsets.

The relationships between initiating stimuli and terminating responses which are to be strengthened need not be isomorphic with the relationships between initiating stimuli and mediating responses. With increasing departures from isomorphism, the acquisition of terminating responses might be retarded, possibly to a degree sufficient to produce some negative transfer. Because the mediating responses might also reduce extralist response interference and stimulus variability, however, the net transfer might still be positive.

If each of the initiating stimuli belongs to two or more response-defined subsets, each stimulus would be expected to evoke two or more different mediating responses, at least during the initial trials. Unless each terminating response is then isomorphic with the two or more responses defining each of the other combinations of subsets, multiple mediating responses to initiating stimuli can be expected to increase generalization among subsets; thus some retardation of the learning of terminating responses would be occasioned. Because of greater trial-to-trial variability in stimulation preceding terminating responses, multiple mediating responses to initiating stimuli may always produce some retardation relative to the maximum positive transfer that is achievable with a single mediating response to each initiating stimulus.

Use of Two-Stage Paradigms

Each of the preceding two-stage paradigms represents a different general case of relationships among the particular type of sets of initiating stimuli, mediating responses and stimuli, and terminating responses. The members of each of these types of sets of initiating stimuli may differ with respect to their complexity, their similarity and other properties, their number, and their probabilities of occurrence (Goss, 1955). Thus, it should be obvious that each particular concept formation task and its attendant conditions requires detailed analysis in terms of presumed stimulus-response elements and initial relationships among these elements, and also in terms of the changes in those elements and relationships which are expected to occur. The nature and strengths of both initial and changed relationships should be expressed as completely and precisely as possible. Association techniques or training controlled by the experimenter are the means of specifying the strengths of initial relationships. It should also be remembered that—except when isolated and controlled—changes in many or all of the stimulus-response relationships present probably occur, if not simultaneously, within rather small blocks of trials.

Finally, some subjects who have formed concepts correctly may not provide verbal reports of the bases for their conceptual behaviors which correspond to experimenters' specifications of the bases for forming particular concepts. When verbal mediating responses do not occur, or occur only partially, sporadically, or during the earlier phases of concept formation, appropriate verbal reports would not be expected. Should the words which constitute subjects' mediating responses differ from the labels preferred by experimenters, subjects' reports might be considered wrong or incomplete. Further, during the course of acquisition the sets of verbal mediating responses may have changed. Should subjects have failed to distinguish such changes or to

indicate when they took place, their reports would seem inaccurate and confused. In addition, if subjects described the terminating responses or both mediating and terminating responses as bases for their conceptual behaviors, while experimenters' specifications were only in terms of verbal mediating responses, subjects' verbal reports would seem unsatisfactory. This would also be the case were the subjects' mediating responses nonverbal.

Differences between the labels used by subjects and those preferred by experimenters can only be determined by careful, detailed analyses of the labeling habits of subjects from a given population. Confusions between mediating and terminating responses can be minimized by ascertaining the temporal sequences of subjects' responses. Only by careful observation of locomotor-manipulative responses (and even this may be inadequate) will it be possible to determine the presence of nonverbal mediating responses.

PLACE OF "ABSTRACT SET OR ATTITUDE," "HYPOTHESES," AND "STRATEGIES" IN THE FRAMEWORK OF THIS ANALYSIS

"Abstract set or attitude" (Goldstein & Scheerer, 1941; Hanfmann & Kasinin, 1942), "hypotheses" (Woodworth, 1958), and "strategies" (Bruner, Goodnow, & Austin, 1956) are terms frequently used to label some aspects of the behavior involved in acquiring concepts as well as to explain success or failure in this process. Unfortunately, these notions have certain features that limit both their experimental usefulness and their explanatory power. In general, their presence or absence is ordinarily determined on the basis of characteristics of the conceptual behaviors observed. Thus, they are post hoc descriptions which usually cannot be used predictively. Often, too, the notions are treated as primary or sole explanations of conceptual behavior when, in fact, other factors—such as types of sets of initiating stimuli, specific attributes of each type of stimuli, and amount and conditions of practice—appear to be of equal or greater importance. And, on the whole, the relationship of these terms to more general theories of behavior is tenuous at best.

In the face of such shortcomings, one course for a stimulus-response analysis of concept formation consists of ignoring the terms entirely. Another course, followed here, is to attempt to assimilate what is meaningful and useful about the terms within the rather rigorous framework that this paper has presented. More specifically, it is suggested that the most meaningful and useful aspects of such notions as abstract set or attitude, hypotheses, and strategies are in part already present in this analysis, and that what is valuable but not present requires only certain translations in order to be assimilated. These aspects are considered below under the following headings: verbal mediating responses and stimuli, strengths of reactions to occurrences and extent of reversal or nonreversal shifts, sequences in which verbal mediating responses occur on single and successive trials, sequences of receptor-orienting responses, and prior habits and persistence of covert or overt verbalization and rehearsal.

Verbal Mediating Responses and Stimuli

Following Fenn and Goss (1957), perhaps the simplest as well as the most common meaning of abstract set or attitude and hypotheses in concept

formation is, conceived narrowly, the occurrence of verbal mediating responses and stimuli. Conceived more broadly, this meaning subsumes the largely pre-experimentally established patterns of relationships: (*a*) between initiating stimuli and mediating responses, where the latter are names for dimensions and values as well as for common elements or variable features, or where they are common responses or meanings that define subsets of initiating stimuli; (*b*) between such mediating responses and other mediating responses; and (*c*) between such mediating responses and terminating responses. The first part of this section on paradigms was largely devoted to an analysis of the role of these patterns of relationships in concept formation. The functional significance for concept formation of this meaning of abstract set or attitude and hypotheses has, therefore, already been considered.

Reactions to Shifts

Also considered earlier were mediating responses that identify shifts in the task. Such responses, it was suggested, should increase in strength with successive reversal or nonreversal shifts and thus mediate changes in whole sets of further mediating responses which name dimensions and values along dimensions. These changes might be called shifts in hypotheses or strategies. And their occurrence could be taken as evidence of the presence of an abstract set or attitude.

Sequences in Which Verbal Mediating Responses Occur

Bruner, Goodnow, and Austin (1956, pp. 81–103, 126–147) have suggested that conditions of presentation of initiating stimuli influence trial-to-trial sequences of choice and "guess" responses. Thus, when subjects could select each successive initiating stimulus, the four sequences or strategies which were distinguished logically were simultaneous scanning, successive scanning, conservative focusing, and focus gambling. When the successive occurrences of initiating stimuli were controlled by the experimenter, they distinguished wholist (whole, focusing) and part-scanning (part) strategies or sequences. Under the condition of presentation in which subjects could select each successive initiating stimulus, the responses which were recorded were terminating responses, first in the form of a choice and then as a guess. Presumably these choices and guesses were preceded, most immediately, by mediating responses consisting of names for the combinations of values of the stimulus selected and the names of the consequent guess of the correct concept. Therefore, the sequences or strategies they distinguished, and which were found to occur to some degree in various subjects, could be regarded as providing some information about sequences of mediating responses through successive selections of stimuli. Under the condition of presentation in which the experimenter determined each successive initiating stimulus, each hypothesis written during the 10-second period following each initiating stimulus could be regarded as congruent with the just preceding mediating response. The sequences of such hypotheses, therefore, probably reflected trial-to-trial sequences of the last mediating response of each trial. No information about intratrial sequences was reported.

The two-stage paradigms of the first part of this section show each initiating stimulus as eliciting only one mediating response which is either the name for a dimension or

value along a dimension, or a single combination of names for dimensions or values. Contingent on both the time subjects have to respond and on subjects' prior experiences with the same or similar stimuli, each initiating stimulus may elicit not one name or combination of names, but a sequence of names or combinations of names. For example, the subject might respond to a particular initiating stimulus with the sequence "red, small, color, size," in which "size" was the last response to occur prior to the appearance of the stimulus eliciting the terminating response to be conditioned to the initiating stimulus. Because of the shorter time interval, the terminating response might be more strongly conditioned to stimuli produced by size than to stimuli produced by the earlier mediating responses. More generally, should terminating responses be most strongly conditioned to stimuli produced by mediating responses which occurred just prior to elicitation of the terminating responses, the sequences with which mediating responses occur and whether those responses are names of values or of dimensions might have marked effects on concept formation.

Whether the effects are facilitative or inhibitory will be contingent on particular conditions. Thus, were size the relevant dimension and large and small the two values along that dimension, the sequence "color, red (or blue), size, large (or small)" should produce greater facilitation than the sequence "size, large (or small), color, red (or blue)." Similarly, sequences in which the last mediating responses were names for common elements of the initiating stimuli rather than names for their variable features should facilitate acquisition of different terminating responses to each subset of initiating stimuli with common features. Inhibitory consequences would be predicted for sequences ending with mediating responses which were names for variable features rather than for common elements. Also, occurrence of the common response to a subset of initiating word stimuli, after more specific associations to those words rather than before such associations, should facilitate; the opposite sequence should inhibit.

A further consideration would be whether the same sequence or different sequences of mediating responses occurred on each presentation of each initiating stimulus or of each member of particular subsets of initiating stimuli. With the sequence, "color, red (or blue), size, large (or small)," for example, constancy of the sequence should be most facilitative, were size the relevant dimension, and most inhibitory, were color the relevant dimension. A reduction in the percentage of times "size, large (or small)" occurred last, and a concomitant increase in the percentage of times "color, red (or blue)" occurred last, should be relatively less facilitative or less inhibitory in the formation of size or color concepts, respectively.

Modes of systematic variation of the components of sequences and of the order in which the components occur can be learned. Therefore, subjects can be expected to differ in the degree to which they have learned to vary the nature and sequences of mediating responses through successive trials. As a result subjects will not only differ with respect to the abstract set or attitude, hypotheses, and strategies with which they began but also with respect to those which are present through successive trials. Whether particular sequences or ways of varying such sequences are facilitative or inhibitory will be contingent on the particular concepts to be formed.

Sequences of Receptor-Orienting Responses

Receptor-orienting responses and their consequences may sometimes be functionally equivalent to mediating responses and stimuli (Goss, 1955). For this reason, abstract set or attitude, hypotheses, or strategies may also be conceived as sequences of receptor-orienting responses.

When initiating stimuli which are relatively small in size are presented at the same place, one at a time, receptor-orienting responses may be of little importance. Possible exceptions are initiating stimuli composed of combinations of common elements or relations and variable features for which the common element or relation of all members of a particular subset have the same location. Should there be some favored point of initial fixation for individual subjects, or for groups of subjects, whether the common element or relation of a particular subset was at that location or at other locations might influence acquisition of the concepts.

Simultaneous presentation of all initiating stimuli or groups thereof, however, might increase the importance of sequences of receptor-orienting responses. Both arrangement of initiating stimuli on the display, and the subjects' pre-experimental and subsequent experimental experiences, should determine the particular sequence of receptor-orienting responses on a given trial. The initial and subsequent fixation points might maximize focusing on successive stimuli whose combinations of values and changes in those combinations were optimal for the formation of particular concepts. If so, such sequences of receptor-orienting responses should facilitate concept formation. For other arrangements of initiating stimuli the same sequences might be inhibitory.

Prior Habits of Verbalization, Rehearsal, and Persistence Therein

Included in Dollard and Miller's (1950, pp. 118–119) set of factors in "social training in the use of higher mental processes" is "training to stop and think." Adolescent and adult subjects explicitly instructed to use verbal mediating responses may differ little, if at all, in the degree to which such responses are activated. However, without such explicit instructions—and therefore largely dependent on the subjects' past experiences with similar tasks—they may or may not stop and think: i.e., they may or may not make overt or covert verbal mediating responses prior to occurrences of terminating responses to the initiating stimuli. Furthermore, some subjects may rehearse such responses between trials while others may think of other things; the latter subjects may in other words, fail to attend to the task continuously. Finally, in the face of initial failures, some subjects may persist in stopping and thinking and in rehearsing while other subjects may temporarily or permanently stop both activities. Up through adolescence the strengths of such habits should be directly related to age. Awaiting detailed determination, however, are both the nature of the relationships of habits of verbalization and rehearsal to age and the effects of such habits on probabilities of occurrence of verbal mediating responses.

In summary, conceived analytically rather than simply as names for certain instructions or for certain changes in terminating responses, the notions of abstract set or attitude, hypotheses, and strategies apparently refer to one or more of the preceding classes of relationships among the stimuli and responses of concept formation tasks. Some of the classes of relationships

include mediating responses and stimuli; those which do not can be expected to have indirect effects on relationships that do involve mediating responses and stimuli.

PRINCIPLES AND PREDICTIONS

Though referred to occasionally—and always assumed—in the preceding section, little direct attention has yet been given to the classes of variables and of general principles involving those variables which enter into explanations of the strengthening, generalization, and weakening of the stimulus-response associations entailed in the one- and two-stage paradigms that have been described. Of obvious relevance are the classes of principles that concern effects on associations or on performance of classes of variables such as: schedules of practice and reinforcement-punishment, the number and both absolute and relative strengths of conflicting responses, the number of stimuli associated with the same response and the strengths of those associations, and the degree of similarity among initiating stimuli and among mediating stimuli. Setting limits to the operation of these classes of variables are the patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses and also conditions of stimulus presentation, such as whether initiating stimuli are presented simultaneously or successively (Bruner et al., 1956) and whether they are all positive, negative, or both positive and negative (Hovland, 1952; Hovland & Weiss, 1953).

It is not the purpose of this paper to make an exhaustive enumeration of the consequences predicted by the application of each class of potentially relevant variables and the principles involving them to the several paradigms or to the various patterns of relationships the paradigms contain. In order to show explicitly how such variables and principles may be profitably combined with the paradigms, however, this final section deals with certain aspects of predictions of the effects of three important classes of variables for which some data are available. These are: strength of associations between initiating stimuli and mediating responses; patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses; and similarity of initiating stimuli. In each case, pertinent experimental studies are described.[6]

Strengths of Associations between Initiating Stimuli and Mediating Responses

The strengths of associations between initiating stimuli and verbal mediating responses will be determined by conditions of practice such as the number and distribution of trials or degree of mastery of those relationships prior to undertaking transfer or criterion tasks. In gen-

[6] Not considered, however, are those studies of the relative effects of reversal and nonreversal shifts which were noted in the first part of the second section. Also ignored are studies (Bensberg, 1958; Carey & Goss, 1957; Fenn & Goss, 1957; Hunter & Ranken, 1956; Wickens & Eckstrand, 1954) which were primarily demonstrations of positive transfer from verbal pretraining to subsequent conceptual behaviors; these demonstrations served as bases for inferences about the functional significance of verbal mediating responses in conceptual behaviors. Several additional experiments (Attneave, 1957; Rhine & Silun, 1958; Shepard & Shaeffer, 1956; Sigel, 1953, 1954; Solley & Messick, 1957; Staats & Staats, 1957; Wulff & Stolurow, 1957) have been excluded because they did not involve either experimentally controlled verbal pretraining or conventional concept formation criterion tasks.

eral, any condition of practice and reinforcement-punishment known to increase or decrease the strengths of stimulus-response associations of multiunit tasks are, through their effects on strengths of associations between initiating stimuli and mediating responses, potential determinants of subsequent performance on transfer or criterion tasks.

In Figures 2 and 7, the relationships between initiating stimuli and mediating responses, and between the initiating stimuli in combination with mediating stimuli and terminating responses, can be described as isomorphic. Put another way, for each different mediating response to a subset of initiating stimuli, there is one and only one terminating response, each of which is different from the terminating response paired with any other mediating response. For such isomorphic patterns of relationships, it is predicted that rate of acquisition of associations between initiating stimuli and terminating responses would be a direct function of strengths of associations between initiating stimuli and mediating responses. Because of generalized responses (errors of generalization, confusions, intralist intrusions), trials to learn associations between initiating stimuli and terminating responses should be related to trials in learning associations between initiating stimuli and mediating responses by an ogival function or by curves showing some slight initial negative transfer rather than being negatively accelerated throughout (Goss, 1955).[7]

Pertinent to this prediction are two recent investigations (Goss & Moylan, 1958; Lacey & Goss, 1959) of the relationship between transfer to conceptual behaviors and strengths of associations between initiating stimuli and presumed verbal mediating responses. In both investigations, the initiating stimuli were 16 blocks, each of which was tall or short, black or white, in combination with top and bottom areas which were large or small, square or circular. In the Goss and Moylan study, nonsense syllable responses or familiar word responses were conditioned to subsets of tall-large, tall-small, short-large, and short-small initiating stimuli. Lacey and Goss used only nonsense syllable responses. The transfer task of both studies was sorting by height-size, and in both the number of blocks sorted by height-size was directly related to degree of mastery of associations between initiating stimuli and presumed mediating responses, as well as to numbers of trials in learning those associations. Unfortunately the resultant curves were not adequate for more precise specification of functions relating direction and degree of transfer to degree of mastery of associations between initiating stimuli and mediating responses, or to trials in learning these associations.[8] As suggested elsewhere for paired-associates learning tasks (Goss, 1955), such specifications are further complicated by the likelihood that the functions are contingent on parameters such as patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminat-

[7] Murdock (1958) argues that with appropriate allowance for generalization responses, the function is negatively accelerated throughout.

[8] An alternative suggestion (Lacey & Goss, 1959) is that greater mastery of experimentally established associations between initiating stimuli and nonsense syllable responses increases the likelihood of arousal of preexperimentally established associations between initiating stimuli and names for dimensions and values along dimensions. Such names might then serve as the actual verbal mediating responses of the transfer or criterion task.

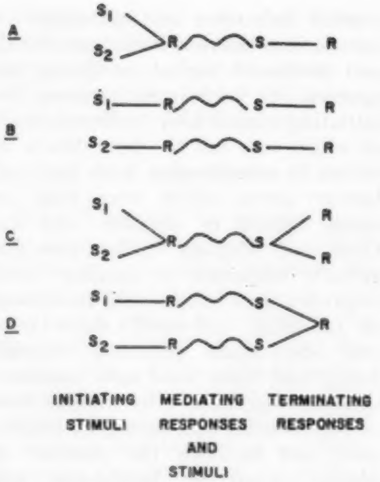

FIG. 8. Four possible extreme patterns of relationships among subsets of initiating stimuli, mediating responses and stimuli, and terminating responses.

ing responses, as well as on the degree of similarity of initiating stimuli.

Patterns of Relationships

Within two-stage paradigms, regardless of the type of sets of initiating stimuli, it is useful to distinguish four extreme patterns of relationships among subsets of initiating stimuli, mediating responses and stimuli, and terminating responses, because each pattern should result in somewhat different conceptual behaviors involving the terminating responses. Figure 8 shows these four patterns. In Patterns *A* and *B*, the relationships between mediating responses and subsets of initiating stimuli are isomorphic with those between terminating responses and subsets of initiating stimuli plus mediating stimuli. This isomorphism does not hold for Patterns *C* and *D*. Pattern *C* is characterized by a common mediating response to both subsets of initiating stimuli and by two terminating responses, one of which is to stimulus compounds consisting of stimuli from the first subset plus the mediating stimulus, and the other of which is to stimulus compounds consisting of stimuli from the second subset plus the mediating stimulus. Pattern *D* is characterized by a different mediating response to each subset of initiating stimuli, and by a common terminating response both to compounds consisting of the stimuli in the first subset plus the stimulus produced by the mediating response to those stimuli, and to compounds consisting of the stimuli in the second subset plus the stimulus produced by the mediating response to those stimuli.

For concept formation tasks involving the relationships of Patterns *A* and *B*, prior acquisition of the associations between subsets of initiating stimuli and a common mediating response should facilitate acquisition of associations between the subsets of initiating stimuli and a common terminating response; prior acquisition of different mediating responses should facilitate acquisition of associations between initiating stimuli subsets and different terminating responses. Greater response-mediated similarity and generalization is the basis for the prediction for Pattern *A*, and greater response-mediated dissimilarity and discrimination is the basis for the prediction for Pattern *B*.

For Pattern *C*, in contrast, the greater similarity of the subsets of initiating stimuli (based on the presence of a common mediating stimulus) should retard acquisition of a different terminating response to each subset. For Pattern *D*, greater dissimilarity of the subsets of initiating stimuli (based on the presence of a different mediating stimulus for the stimuli of each subset) should slow the learning

of a common terminating response to each subset.

Lacey (1959) tested each of these predictions. His stimuli were eight line drawings of faces or houses, each of which was printed on pink, light blue, light yellow, and light green paper. Eight to 11 year old children first learned either a common or different nonsense syllable mediating response to two subsets of initiating stimuli. The transfer or criterion task was acquisition of a new set of nonsense syllable responses as either a common or different terminating response to those same subsets of initiating stimuli. Thus the relationships of terminating responses to initiating stimuli and to mediating responses and stimuli were those of Patterns *A*, *B*, *C*, or *D*.

Measured against the performance of control groups, whose prior training controlled for facilitation due to warm up and receptor-orienting responses, positive transfer was obtained with Patterns *A* and *B* and negative transfer occurred with Patterns *C* and *D*. Therefore, as predicted, the pattern of relationships among subsets of initiating stimuli, mediating responses and stimuli, and terminating responses determined whether positive or negative transfer occurred. Lacey's results also suggested that Pattern *B* might produce greater relative positive transfer, though no greater absolute positive transfer, than Pattern *A*. Patterns *C* and *D*, however, did not seem to differ with respect to either relative or absolute amounts of negative transfer.

Similarity of Initiating Stimuli

For Patterns *A*, *B*, *C*, and *D*, similarity of sets of initiating stimuli might influence amount and perhaps direction of transfer from verbal pretraining to subsequent conceptual behaviors. For Patterns *A* and *D*, disregarding mediating responses and stimuli, similarity within and between subsets of initiating stimuli should be directly related to ease of learning associations between those stimuli and a common terminating response. Patterns *B* and *C* involve acquisition of discriminative terminating responses to initiating stimuli. Rate of acquisition of those associations should be directly related to similarity of stimuli within subsets of initiating stimuli and inversely related to similarity between those subsets. When verbal mediating responses and stimuli are considered, similarity of initiating stimuli might modify the expected positive transfer with Patterns *A* and *B* and the expected negative transfer with Patterns *C* and *D*.

Present data and theory do not warrant attempts to develop more exact predictions of the influence of similarity of initiating stimuli on direction and amount of transfer. However, since the second variable of Lacey's (1959) experiment was two degrees of similarity of the members of the sets of face and house stimuli, some pertinent data are available.

For Patterns *A* and *D* together, disregarding verbal mediating responses and stimuli, similarity was directly related to mastery of associations between initiating stimuli and terminating responses; inverse relationships were obtained with Patterns *B* and *C* together. For Pattern *A*, while absolute amount of positive transfer was directly related to similarity, an inverse relationship was obtained for relative amount of transfer. For Pattern *D*, both absolute and relative amounts of positive transfer were inversely related to similarity. For Patterns *B* and *C*, both absolute and relative amounts of positive transfer were directly related to similarity. However, since most of the relation-

ships for each pattern separately were not statistically significant, at best they provide hypotheses for replicatory investigations.

In general, for concept formation tasks involving prior strengthening of presumed mediating responses, the findings presently available suggest that conceptual behaviors involving terminating responses are influenced by: strengths of associations between initiating stimuli and mediating responses; patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses; and similarity of initiating stimuli. Furthermore, these findings are reasonably consistent with predictions based on two-stage paradigms in combination with principles of the role of these and other classes of variables in the strengthening, generalization, and weakening of stimulus-response associations.

SUMMARY

The purpose of this paper was to analyze the role of verbal mediating responses in concept formation. First summarized was the historical development of stimulus-response analyses of conceptual behaviors which have emphasized the role of mediating responses and stimuli, particularly verbal mediating responses. The influence of Max Meyer and Watson on the behavioristic analyses of the 1920s was noted. Although Birge, Miller and Dollard, Cofer and Foley, and Gibson furthered such analyses in the early 1940s, only the more detailed recent analyses of Baum, Osgood, Mandler, Goss, Kendler, and others have led to hypotheses which have been tested experimentally.

The first section provided a general specification of concept formation tasks and described the relationship between concept formation and conventional paired-associates tasks. The second section first described the structures and some explanatory consequences of one-stage and two-stage paradigms of conceptual behaviors with each of three types of sets of initiating stimuli. Some precautions in the use of these paradigms were then noted, and assimilation within the present analysis of the notions of abstract set or attitude, hypotheses, and strategies was proposed. The third section showed the complementary relationship between the one-stage and two-stage paradigms and classes of variables and principles involving those variables which enter into explanations of the strengthening, generalization, and weakening of the component stimulus-response associations. Two-stage paradigms in combination with some of these principles were then used to generate sample predictions of effects on concept formation of: strengths of relationships between initiating stimuli and mediating responses; some patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses; and relative similarity of initiating stimuli.

REFERENCES

ATTNEAVE, F. Transfer of experience with a class-schema to identification-learning of patterns and shapes. *J. exp. Psychol.*, 1957, **54**, 81–88.

BAUM, M. H. A study in concept attainment and verbal learning. Unpublished PhD dissertation, Yale University, 1951.

BENSBERG, G. J., JR. Concept learning in mental defectives as a function of appropriate and inappropriate "attention sets." *J. educ. Psychol.*, 1958, **49**, 137–143.

BIRGE, J. S. The role of verbal responses in transfer. Unpublished PhD dissertation, Yale University, 1941.

BRUNNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

BUSS, A. H. Reversal and nonreversal shifts in concept formation with partial reinforce-

ment eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

CAREY, J. E., & GOSS, A. E. The role of verbal labeling in the conceptual sorting behavior of children. *J. genet. Psychol.*, 1957, **90**, 69–74.

COFER, C. N., & FOLEY, J. P., JR. Mediated generalizations and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.

DASHIELL, J. F. *Fundamentals of objective psychology.* Boston: Houghton Mifflin, 1928.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

ELKONIN, D. B. The physiology of higher nervous activity and child psychology. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957.

FEDERER, W. T. *Experimental design: Theory and application.* New York: Macmillan, 1955.

FENN, J. D., & GOSS, A. E. The role of mediating verbal responses in the conceptual sorting behavior of normals and schizophrenics. *J. genet. Psychol.*, 1957, **90**, 59–67.

FREEDMAN, J. L., & MEDNICK, S. A. Ease of attainment of concepts as a function of response dominance variance. *J. exp. Psychol.*, 1958, **55**, 463–466.

GELFAND, S. Effects of prior associations and task complexity upon the identification of concepts. *Psychol. Rep.*, 1958, **4**, 568–574.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

GORMEZANO, I., & GRANT, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.

GOSS, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62**, 20–31.

GOSS, A. E. University of Massachusetts conference on problem solving. Amherst, Massachusetts, June 19–21, 1956.

GOSS, A. E., & GREENFELD, N. Transfer to a motor task as influenced by conditions and degree of prior discrimination training. *J. exp. Psychol.*, 1958, **55**, 258–269.

GOSS, A. E., & MOYLAN, M. C. Conceptual block-sorting as a function of type and degree of mastery of discriminative verbal responses. *J. genet. Psychol.*, 1958, **93**, 191–198.

GRAY, J. S. A behavioristic interpretation of concept formation. *Psychol. Rev.*, 1931, **38**, 65–72.

GRIFFITH, B. C., & SPITZ, H. H. Some relationships between abstraction and word meaning in retarded adolescents. *Amer. J. ment. Defic.*, 1958, **63**, 247–251.

GRIFFITH, B. C., SPITZ, H. H., & LIPMAN, R. S. Verbal mediation and concept formation in retarded and normal subjects. *J. exp. Psychol.*, 1959, **58**, 247–251.

HANFMANN, E., & KASININ, J. Conceptual thinking in schizophrenia. *Nerv. ment. dis. Monogr.*, 1942, No. 67.

HARROW, M., & FRIEDMAN, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–598.

HEIDBREDER, E. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **35**, 173–189. (a)

HEIDBREDER, E. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35**, 191–223. (b)

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **59**, 461–472.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

HULL, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**(1, Whole No. 123).

HUNTER, G. F., & RANKEN, H. B. Mediating effects of labeling on sorting behavior and judgments of similarity. Paper presented at Eastern Psychological Association, Atlantic City, March 1956.

KELLEHER, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **48**, 165–174.

KENDLER, H. H., & KENDLER, T. S. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

KENDLER, H. H., & MAYZNER, M. S., JR. Reversal and nonreversal shifts in card-sorting tests with two and four categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

LACEY, H. Mediating verbal responses and stimulus similarity as factors in conceptual naming by school-age children. Unpublished PhD dissertation, University of Massachusetts, 1959.

LACEY, H., & GOSS, A. E. Conceptual block sorting as a function of number, pattern of assignment, and strength of labeling responses. *J. genet. Psychol.*, 1959, **94**, 221–232.

LIUBLINSKAYA, A. A. Development of children's speech and thought. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

METZGER, R. A comparison between rote learning and concept formation. *J. exp. Psychol.*, 1958, **56**, 226–231.

MEYER, M. F. *The fundamental laws of human behavior.* Boston: Gorham, 1911.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

MURDOCK, B. B., JR. Intralist generalization in paired-associate learning. *Psychol. Rev.*, 1958, **65**, 306–314.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

REED, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.*, 1946, **36**, 71–87.

RHINE, R. J., & SILUN, B. A. Acquisition and change of a concept attitude as a function of consistency of reinforcement. *J. exp. Psychol.*, 1958, **55**, 524–529.

RICHARDSON, J. The relationship of stimulus similarity and numbers of responses. *J. exp. Psychol.*, 1958, **56**, 478–484.

SHEPARD, W. O., & SHAEFFER, M. The effect of concept knowledge on discrimination learning. *Child Develpm.*, 1956, **26**, 173–178.

SIGEL, I. Developmental trends in the abstraction ability of children. *Child Develpm.*, 1953, **24**, 131–144.

SIGEL, I. The dominance of meaning. *J. genet. Psychol.*, 1954, **85**, 201–207.

SIMMONS, A. J., & GOSS, A. E. Animistic responses as a function of sentence contexts and instructions. *J. genet. Psychol.*, 1957, **91**, 181–189.

SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**(4, Whole No. 191).

SOLLEY, C. M., & MESSICK, S. J. Probability, learning, the statistical structure of concepts, and the measurement of meaning. *Amer. J. Psychol.*, 1957, **70**, 161–173.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54**, 74–80.

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229–238.

WATSON, J. B. Is thinking merely the action of language mechanisms? Part V. *Brit. J. Psychol.*, 1920, **11**, 87–104.

WEISS, A. P. *A theoretical basis of human behavior.* Columbus, Ohio: Adams, 1925.

WICKENS, D. D. Stimulus-response theory as applied to perception. In, *Kentucky symposium: Learning theory, personality theory, and clinical research.* New York: Wiley, 1954.

WICKENS, D. D., & ECKSTRAND, G. A. Transfer of perceptual set. *J. exp. Psychol.*, 1954, **47**, 274–278.

WOODWORTH, R. S. *Dynamics of behavior.* New York: Holt, 1958.

WULFF, J. J., & STOLUROW, L. M. The role of class-descriptive cues in paired-associates learning. *J. exp. Psychol.*, 1957, **53**, 199–206.

(Received December 9, 1959)

Psychological Review
1961, Vol. 68, No. 4, 275-284

COSTS AND PAYOFFS ARE INSTRUCTIONS [1]

WARD EDWARDS

Engineering Psychology Group, University of Michigan

This paper states the problem of internally contradictory or ambiguous instructions; shows how the specification of costs, payoffs, and exchange rates solves it; and discusses necessary properties of adequate solutions.

First I must distinguish three different purposes for which costs and payoffs (in money or any other valuable commodity, real or imaginary) might be used in psychological experiments on human subjects. Suppose that you pay the subject by the hour for his time, and that your experiment takes a fixed amount of time. Then your payment serves only as a motivator. Suppose that you ask the subject to memorize a list of nonsense syllables and pay him a dime for every one he remembers correctly. In fact, the only way he knows that any remembered syllable is correct is that you hand him a dime as soon as he says it. Then your payments, in addition to whatever motivating function they may serve, also serve as information givers. Finally, suppose that in the nonsense syllable memorization experiment you tell the subject that he will be charged a dime for every syllable he "remembers" which was not in fact on the original list. Now your payments, in addition to serving motivational and informative purposes, also serve as instructions, since they tell the subject exactly the relative desirability or undesirability of correct syllables, incorrect syllables, and omissions. If you withheld all payments and all other information about how the subject was doing until the experiment was over, your payments (or, more precisely, your statements about them before the experiment) would still serve instructional as well as motivational purposes, though they would no longer serve as informative feedback.

This paper is concerned only with the instructional function of costs and payoffs. Many recent experiments have used costs and payoffs primarily for instructional purposes (e.g., Becker, 1958; Edwards, 1956; Goodnow & Pettigrew, 1956; Irwin & Smith, 1957; Lawrence & LaBerge, 1956; Lewis & Duncan, 1956, 1957, 1958; Siegel & Goldstein, 1959; Tanner & Swets, 1954).

Ambiguous and Internally Contradictory Instructions

Many of the instructions most commonly used in psychological experiments are at best ambiguous and at worst internally contradictory. For example, consider a speeded intelligence test. Its instructions say: "Answer as many questions as you can. You have 10 minutes for this part of the test." What is the subject supposed to do? Should he make certain that each answer is correct, thus minimizing errors but dealing with relatively few questions? Should he answer as many questions as possible, guessing when he does not know the answer? Or should he adopt some compromise between these strategies; and if so, what compromise? The instructions do not say. In fact, the instructions tell him to perform an im-

[1] This work was conducted by Project MICHIGAN under Department of the Army Contract DA-36-039 SC-78801, administered by the United States Army Signal Corps.

possibility; they say that he should simultaneously maximize the number of questions answered and minimize the number of errors. These instructions are internally inconsistent. A computing machine would reject as insoluble a problem presented with such instructions. Human beings, more tractable and less logical, perform such tasks every day. The only way they can do so is to provide some kind of self-instruction which supersedes the impossible instructions.

The problem is not confined to testing situations; it arises in any experiment which includes two or more inconsistent measures of quality of performance. Three measures which lend themselves particularly well to this kind of inconsistency are time, number of correct responses, and number of incorrect responses. These measures are rather often used in psychological experiments. As these measures suggest, the problem of inconsistent or ambiguous instructions is most likely to arise when perfect performance is specified as ideal (e.g., all questions should be answered correctly) but no information is provided which would enable the subject to evaluate the relative undesirability of various kinds of deviations from perfection.

THE MATHEMATICAL NATURE OF THE PROBLEM

The problem I am raising is essentially a mathematical one. Many psychological experiments, and indeed just about all those which deal with the more "volitional," less reflexly determined kinds of human behavior, use instructions which contain the phrase "Do the best you can" or its equivalent. (This paper is not relevant to experiments in which the subject is not given some such instruction.) This, in mathematical language, is an instruction to maximize or minimize some function whose nature is determined by the experimental situation. So long as only one function is to be maximized or minimized, such instructions present no problem. If, for example, a subject is shown a nonsense figure, and then later shown a group of four such figures and required to point to the one most like the original, an instruction to maximize correct answers creates no ambiguity.

But as soon as another function which the subject is also supposed to maximize or minimize is introduced into the same situation, the problem of possible inconsistency arises. It is only rarely and by coincidence in mathematics that two functions will have the same maxima and/or minima. If they do not, then the subject cannot do "the best he can" with respect to both functions simultaneously. Suppose we modify the pattern recognition experiment of the previous paragraph by requiring a judgment of "same" or "different" for each of the four figures which may be like the original. To maximize the probability of a correct identification, the subject should say "same" in response to each of the four figures. To minimize the number of incorrect positive identifications, he should call them all "different." Various different weightings of the relative importances of true positives, false positives, true negatives, false negatives, and failures to respond will lead to a number of different optimal strategies. Unless the experimenter provides some basis for such differential weights, no optimal behavior pattern can be specified. Typical instructions for such an experiment might direct the subject to minimize false positives, false negatives, and failures to respond while maximizing true positives and true negatives. Such instructions are, of course, impossible to carry out unless the subject can produce a perfect

performance (a possibility which this paper will ignore). This example requires not two extreme values, but five, of which three are independent of one another. Please note, however, that very simple situations may contain the same ambiguity. Consider a simple psychophysical experiment in which the subject is asked whether or not a tone was presented on a given trial. The implicit instruction is to maximize the number of statements that a tone was present when it in fact was and to minimize that number when in fact it was not; this instruction is internally inconsistent.

The problem is complicated further by a distinction, clear-cut in mathematics but not in the application of mathematics to experiments, between extreme values and boundary values. A boundary value is simply a value of a variable which that variable cannot exceed (in the case of an upper boundary) or be less than (in the case of a lower boundary). If, for instance, a task has a time limit, but no one cares how long it takes to perform the task so long as it is within the time limit, that time limit is a boundary value, not an extreme value. It is not mathematically self-contradictory to ask a subject to maximize or minimize a given function (e.g., number of correct answers) within given boundary values on other functions (e.g., a time limit and a limit on the length of time the subject may look at the stimulus). The point is that it must be impossible for the subject to go outside the boundary value. If, for example, the problem has a time limit and the subject exceeds it, then to score that trial as a failure, or indeed to score it in any other way, converts the time limit from a boundary value into an extreme value on a dichotomized scale, and thereby raises once again the problem of inconsistent extreme values on different dimensions. To avoid raising this problem, the experimenter must prevent the occurrence of trials on which the subject exceeds the time limit.

PAYOFF MATRICES AND EXCHANGE RATES

Any instruction of the general form "Do the best you can" implies the existence of at least one method for evaluating the quality of what the subject does, that is, a definition of the word "best." I shall call this evaluation method a *measure of value*. The difficulty with which this paper is concerned is that many experiments offer a number of inconsistent measures of value. This phrasing of the problem implies the obvious solution: define a single measure of value for any given experiment, and then instruct the subject to maximize it. (Of course other instructions concerning the nature of the apparatus and similar matters must also be given, and understood.) I shall call this single measure a final value measure, since it is usually constructed by combining other more primitive value measures.

The solution given above requires two successive steps. The first is the definition of a payoff matrix. A payoff matrix is simply a rectangular matrix with the various courses of action available to the subject on one dimension (usually the vertical one) and the various possible states of the world which determine the outcome of each course of action on the other dimension. For example, consider the pattern recognition experiment which has already been discussed in the version in which the subject must choose one and only one of the test patterns as most similar to the original pattern. Table 1 gives a payoff matrix appropriate for that experiment. Note that only two values, good and bad, appear

TABLE 1
HYPOTHETICAL PAYOFF MATRIX FOR PATTERN RECOGNITION EXPERIMENT

Available courses of action: subject chooses	Possible states of the world—correct pattern is:			
	1	2	3	4
1	good	bad	bad	bad
2	bad	good	bad	bad
3	bad	bad	good	bad
4	bad	bad	bad	good

in the matrix. So long as all wrong answers are equally bad and all right answers equally good, the strategy which maximizes the subject's expected payoff is easy to state: he should choose the answer he considers the most probably correct. (If two or more answers have equal and highest probability he may choose among them any way he likes.) Thus by this particular choice of payoff matrix the subject's optimal strategy by a particular but appropriate definition of optimality has been defined so that his responses will give the best information he is able to give about the degree to which he recognizes a pattern. If other information (e.g., information about the degree of similarity between the original pattern and the test patterns) is needed, other payoff matrices would be appropriate.

The preceding example illustrates the fact that a final value measure need not be stated in numerical terms. In situations (very rare in psychological experiments) in which each course of action available to the subject has precisely one possible outcome, an ordinal measure of value is always sufficient. In situations in which the subject is uncertain about the outcome of at least one course of action, a final value measure defined on at least an interval scale is usually necessary (for an explanation of this statement, see Edwards, 1954b). On an interval scale, two points may be assigned arbitrarily, after which all others are fixed. This means that there is a big difference between situations in which only two levels of the final value measure are necessary and situations which require more than two levels. If three or more levels on the final value measure are necessary, the spacing of these levels must be specified. It is often convenient to choose one interval as the unit and express the other intervals in terms of it.

What about situations in which several different value dimensions contribute to the final value measure? A separate payoff matrix may be prepared for each different value dimension. It is possible that all these separate payoff matrices may have the same characteristics, so that any act which maximizes one value dimension will automatically maximize all others. But this is extremely unlikely. More often, as the examples at the beginning of this paper illustrated, actions which maximize one dimension will not maximize others; I shall call any two such dimensions *inconsistent*. In order to obtain a single final value measure (to be entered in a single payoff matrix) from a set of inconsistent value dimensions, what I shall call *exchange rates* must be established. An exchange rate, as the term is used here, is simply a function defining the relationship between two different measures of value. For example, the custom of scoring four-alternative multiple-choice examination questions (or four-alternative forced-choice psychophysical responses with omissions permitted) by counting the number of right answers and substracting one-third the number of wrong answers

defines two exchange rates, one between right and wrong answers and the other between right answers and failures to answer.

It is not necessary to define exchange rates formally in order to specify final value measure entries in a payoff matrix. This statement, though true, is misleading. If you have a payoff matrix completely filled in with final value measures for a situation including two or more inconsistent value dimensions, it automatically gives as much information as is necessary about exchange rates for all values of interest in the particular situation, though not for any other values. You can start with the exchange rates or with the payoff matrix; it amounts to the same thing, except that the exchange rates generalize more easily to new situations. There is one notable exception to the assertion that the experimenter must specify exchange rates if the experiment requires the subject to cope with inconsistent value dimensions. The exception is experiments which are specifically designed to discover the subject's natural exchange rate between inconsistent value dimensions. Rarely can an empirical exchange rate be determined incidentally, however, in the course of an experiment primarily designed to measure other things.

No necessary relationship exists between the number of value dimensions or the number of levels available on each and the number of levels of the final value measure. Obviously the maximum possible number of levels of the final value measure is the product of the numbers of levels of the various value dimensions. A final value measure which has fewer levels than that in effect asserts that an exact equivalence exists between two or more different combinations of levels of the various value dimensions; such an assertion is entirely appropriate if the experimenter intends that such an equivalence should exist.

A Criterion for Adequate Instructions

So far this paper has asserted that instructions will be free from internal inconsistency and ambiguity if they include a payoff matrix defined in terms of a final value measure, which in turn must usually be defined by means of a set of exchange rates. How can the experimenter be sure that a set of instructions has this desirable property? A very simple check exists. A sufficient criterion for consistent and unambiguous instructions is this: if, in addition to the instructions which the subject is actually given, he were told the probabilities of each possible outcome of each course of action available to him, he would then have enough information to be able to select an optimal course of action. By "optimal" I mean a course of action which maximizes the expected value of the final value measure; except in unusual cases no other definition of optimality is appropriate when probabilities of outcomes are assumed known. In other words, if the subject has all the information about costs and payoffs (though not about probabilities) that he would need to select a strategy which maximizes expected value, he has been unambiguously instructed. This means that the experimenter can always protect himself from ambiguity and inconsistency simply by asking himself: "Does the subject have an optimal strategy available to him and, if he knew the probabilities of the various possible outcomes, would he have enough information to figure out what it is?" If the answer is yes, all is well; if the answer is no, the instructions should be examined suspiciously for ambiguity and internal contradictions.

Of course the preceding criterion carries no implication whatever that the subject will in fact adopt his optimal strategy. For one thing, he ordinarily doesn't have (at least to start with) the information about probabilities which he would need to identify it. For another thing, plenty of evidence exists that subjects do not in fact maximize expected value even when they could do so easily (see Edwards, 1954b). The criterion is exclusively concerned with the experimenter's instructions, not with the subject's behavior.

How can the experimenter apply this criterion? He must first, of course, know himself what final value measure payoff matrix he wants the subject to work with. Then he must communicate the information to the subject in such a way that the subject understands it. This can be done by verbal or written instruction prior to the experiment, by experience during training trials or during the experiment, or by some combination of the two. Of course experience, either in training trials or during the experiment, serves the additional function of giving the subject information about the probabilities of each of the possible outcomes with which he may be confronted. Payoffs which occur during the experiment thus may serve two simultaneous functions: they may instruct the subject about the relevant payoff matrix, and they may inform him about relevant probabilities. In animal experiments these two separate functions are customarily performed simultaneously by the same payoffs. One of the great advantages of human over animal subjects is the possibility of performing these two functions separately. In principle, of course, it would be possible to instruct men in advance concerning the probabilities they will encounter and allow them to induce the relevant payoff matrices from experience during the experiment; I know of no such experiments.

I have said that the proposed criterion of adequate instruction is sufficient, rather than necessary; this implies that exceptions can be found in which the criterion is not satisfied but the instructions are nevertheless unambiguous. What are these exceptions? They are the cases in which the optimal strategy for any one outcome is identical with that for all other possible outcomes, and so only ordinal information about payoffs is necessary to select the optimal strategy. In practice, such a situation would arise in an experiment on the effect of an irrelevant stimulus on some kind of performance. Distraction experiments or experiments on the effect of genuinely redundant information fall in this category. But even in such experiments the presence of other uncertainties may make the proposed criterion necessary as well as sufficient; that would be the case, for example, in an experiment on the effect of irrelevant information on concept identification if the subject had the options of several alternative responses or of not responding. The presence or absence of the irrelevant information on a particular trial would make no difference to his optimal strategy and so would not require interval scale payoff information, but his uncertainty about the correctness of his responses would make such information necessary for selection of an optimal strategy—and so for unambiguous instruction.

So What?

I have asserted that many experiments and tests include ambiguous or self-contradictory instructions, that the problem arises because of inconsistent value dimensions, that the way to avoid

it is to provide a single final value measure, and that the nature of the inconsistent value dimensions and of the problem determine what kind of final value measure is needed. I believe that the statements summarized above are logical, not empirical, and so not controversial—probably an instance of absurd optimism. But in any case I have not yet dealt with the crucial question, which is: so what? What is wrong with internally contradictory instructions? Why should the experimenter be able to specify an optimal strategy for the subject to use, especially when the experimenter knows that no one will use it?

These questions get at the heart of the matter. Obviously there is no good reason why instructions should not contain internal contradictions, and no good reason why the experimenter should be able to specify an optimal strategy, unless these properties make a difference in the value of the experiment or the test. A change in an experimental procedure is desirable if it reduces experimental error, or makes the experiment easier to interpret or otherwise more meaningful. I propose the radical hypothesis that the removal of internal contradictions from instructions has both of these advantages.

The reason why internal contradictions increase experimental error is obvious. If the instructions as given are impossible to carry out, the subject must, in order to do anything at all, somehow supplement or change them from within himself. If all subjects supplement their instructions in the same way, then the only problem the experimenter faces is that of figuring out what the supplementary instructions are—but this is an unlikely outcome, in view of the ubiquity of individual differences. More probably each subject will supplement his instructions differently from all others. Furthermore, he may change his self-instruction in the middle of the experiment. If each subject is operating under instructions different from those used by all other subjects, it is only natural to expect this fact to produce larger individual differences than might otherwise be the case. These individual differences can be explained only if the experimenter discovers what each subject conceives the task to be, which is usually impossible. Otherwise, these individual differences will contribute to experimental error.

The reason why internal contradictions in the instructions make experiments harder to interpret is less obvious, but no less clear-cut. This whole discussion has assumed that subjects are given evaluative instructions: "Do your best." Almost all data based on evaluative instructions are analyzed against evaluative criteria: errors, time, correct answers, or other similar criteria. But such an analysis is far less likely to be psychologically meaningful if the subject had in mind evaluative criteria partially or wholly different from those the experimenter uses. If you give a psychology examination, instruct the examinees to answer as many questions as possible, and then grade them on their grammar, it seems unlikely that you will get very useful information, either about their grammatical skills or their psychological knowledge.

This paper began with the statement that money rewards may serve a primarily instructional function. By now, the argument is obvious. Money is probably the most universally used and understood evaluative dimension in our culture; almost all subjects will understand the statement: "Your purpose in this experiment is to go home with as much money as possible." Furthermore, it is probably easier to re-

duce other value dimensions to money than to any other single dimension. This is particularly true of experiments on problem solving, conceptualizing, and decision making, in which the intent of the subject is always of crucial importance in understanding what he did.

This does not imply that real money must change hands whenever money serves as the final value dimension. Experiments (e.g., Edwards, 1954a) have shown that subjects will respond in much the same way to imaginary money as to real money. An example which uses only imaginary money is a recent experiment by Lawrence and LaBerge (1956). They required their subjects to record three variables of a set of stimuli presented tachistoscopically. One group was told to imagine that all three dimensions were equally important, and that they were being paid $34 for each correct report. Another group was told to imagine that they were being paid $100 for each correct report about one specified dimension, but only $1 for each correct report of either of the other two dimensions. A third group was told to imagine that they were being paid $100 for each correct report about one specified dimension, but nothing for any other report. As might be expected, the results of the last two groups were highly similar to each other and highly dissimilar to the results of the first group. This is a fine example of how money as a final value measure can be used exclusively for instructional, rather than motivational purposes.

An explicitly defined payoff matrix is enough to ensure clear instructions (the sole topic of this paper). But the experimenter could choose any one of a very large number of explicit payoff matrices. Which one he in fact chooses will have an enormous effect on the results he gets. How should he go about deciding among the many possibilities? Obviously it depends on what he wants to find out; no general rule is possible. In fact the judicious selection of appropriate payoff matrices is becoming a major part of the art of experimental design, and instances exist in which weeks have been devoted to the meticulous design and pretesting of a payoff matrix which will do exactly what it is supposed to. Even the methods of psychophysics are being changed as a result of recognition that the choice of a payoff matrix makes a big difference to psychophysical results; Tanner and Swets (1954) have exhibited that changing the values entered in the payoff matrix for an auditory threshold experiment changes the "threshold" (a concept they prefer not to use) found in an orderly and systematic way predictable from a theory which asserts that the subject is attempting to maximize the expected value of his final value measure. The success or failure of such maximization models is irrelevant to the criterion for clear instructions which is the subject of this paper, but rise in interest in such models both in psychophysics and in behavioral decision theory adds to the theoretical and experimental importance of systematic manipulation of payoffs as independent variables.

Of course, money is not the only available final value measure. Errors, time, points, effort—almost anything which can be counted or measured can be a final value measure. Furthermore, there is no necessity that the final value measure be a (theoretically) infinitely divisible entity like money. That is why the preceding discussion placed so much emphasis on the number of levels of the final value measure needed; in cases in which only a few (ideally, two) levels are needed it is usually easy to tell the subject what he should

do without resorting to some external variety of instruction. But money, though not the only available final value measure, is certainly the best. It is unambiguous, it is tangible, it is infinitely divisible; and so will fit almost any problem; and it has useful motivating properties in addition to being a vehicle for instructions.

CONCLUSION

This paper makes a recommendation about experimental design. Unless self-instruction or response to internally contradictory instructions is the problem to be studied, experiments should be designed so that each subject has enough information to resolve ambiguities about how to evaluate the consequences of his own behavior which are inherent in conflicting value dimensions. That means that the subject should have the information about costs and payoffs (though not about probabilities) necessary to evaluate each course of action relative to all others available to him; in other words, if the subject were to be told the probabilities of each possible outcome of each course of action available to him, he should have enough information so that he could then identify unambiguously an optimal strategy (optimal in the sense of maximizing expected value of the final value measure). When an experiment necessarily involves conflicting value dimensions, the conflict can be resolved by specifying exchange rates among the value dimensions which reduce them to a single final value measure. Although money is not the only such measure available, it is very often the best.

SUMMARY

This paper considers the problem of internally contradictory or ambiguous instructions to subjects in psychological experiments. Such instructions often result when subjects must maximize or minimize several dependent variables simultaneously. Time, correct answers, and errors are variables which often are included in internally contradictory instructions. Such contradictions can be resolved by specifying enough information so that subjects have an optimal strategy open to them and, if they knew the probabilities of the various outcomes of each possible course of action, could know what that optimal strategy is. Such specification will usually require that subjects be informed about the exchange rate among the various value dimensions present in the experiment. An exchange rate reduces many value dimensions to one. The most commonly useful value dimension in psychological experiments is money.

REFERENCES

BECKER, G. M. Sequential decision making: Wald's model and estimates of parameters. *J. exp. Psychol.*, 1958, **55**, 628–636.

EDWARDS, W. Probability preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, **67**, 56–67. (a)

EDWARDS, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380-417. (b)

EDWARDS, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

GOODNOW, J. J., & PETTIGREW, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51**, 385–392.

IRWIN, F. W., & SMITH, W. A. S. Value, cost, and information as determiners of decision. *J. exp. Psychol.*, 1957, **54**, 229–232.

LAWRENCE, D. H., & LABERGE, D. L. Relationship between recognition accuracy and order of reporting stimulus dimensions. *J. exp. Psychol.*, 1956, **51**, 12–18.

LEWIS, D. J., & DUNCAN, C. P. Effect of different percentages of money reward on extinction of a lever-pulling response. *J. exp. Psychol.*, 1956, **52**, 23-27.

LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as functions of percent-

age of reinforcement and amount of reward. *J. exp. Psychol.*, 1957, **54**, 115–120.

LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55**, 121–128.

SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37-42.

TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

(Received December 30, 1959)

Psychological Review
1961, Vol. 68, No. 4, 285-288

THEORETICAL NOTES

COMMENT: THE DISTINCTIVENESS OF STIMULI

EDWARD A. FEIGENBAUM AND HERBERT A. SIMON
Carnegie Institute of Technology

In his paper with the above title, Bennet B. Murdock, Jr. (1960) makes some quantitative predictions of the shape of the bowed serial position curve:

> Ideally the applicability of the *D* scale to the serial position curve would be determined by comparing its predictive power to that of other theories. Unfortunately, however, no other theories have been worked out in sufficient detail to permit quantitative predictions to be made (p. 26).

As Murdock himself points out, such predictions have been made with Hullian theory (Hull, Hovland, Ross, Hall, Perkins, & Fitch, 1940)—although not a priori. They have also been made by Atkinson (1957) with what might be described as a stochastic version of Hullian theory. In both cases, good fits are achieved, but the convincingness of the results is diminished by the fact that several parameters are available for adjustment of the fit.

The authors of this note have, however, proposed an alternative theory which, like Murdock's contains no adjustment parameters, and hence can be subjected to rather severe test in fiitting it to the observed data. Unfortunately, the theory has only been described thus far in an unpublished thesis (Feigenbaum, 1959) and an equally unpublished dittoed memorandum (Feigenbaum & Simon, 1958) partly because a paper on it was rejected a year ago by the Editor of the *Psychological Review*. (We are mentioning, not contesting, the action.)

Our purpose in this note is to give some quantitative predictions from our theory that may be compared with Murdock's predictions and with data, so that the relative predictive powers of the theories may be tested. We shall refrain from comment on the structure of any of the theories except insofar as this is relevant to the prediction problem.

Our theory, like Hull's and Atkinson's, predicts the shape of the serial position curve in terms of percentage of total errors at each serial position instead of percentage of total correct responses at each position—the measure Murdock uses. As Murdock observes, the shape of the curve in terms of correct responses will vary with degree of learning. This is known to be empirically true (see, for example, Hull et al., 1940, p. 183). It is also known empirically, and can be predicted from our theory, that the shape of the curves will vary, but only slightly, with such parameters as difficulity of list and presentation rate. In order to use any of these theories to predict the correct-response curve, we must specify the criterion to which the list is learned, and the average number of trials to criterion. The predictions that follow apply to lists that are learned to the criterion of one correct anticipation and that are of such difficulty that the number of trials to criterion equals the number of syllables. Quantitatively, the latter condition is not very important, but the former is.[1]

Briefly, our theory is that the bowing of the curve is a consequence of the strategy the subject adopts (consciously or not) in the learning task. In the absence of contrary instructions from the experimenter or striking inhomogeneities in the list, a subject will first learn the initial list items, then will work from the two ends of the list as "anchor

[1] Murdock's theory only predicts for one degree of learning and does not specify what degree this is. Our theory, like Hull's and Atkinson's, makes specific predictions for *any* criterion and predicts that the bowing of the curve will decrease with degree of learning.

TABLE 1

PREDICTED SERIAL POSITION CURVES FOR LISTS OF 8–15 ITEMS
("Anchor Point" Theory)

Serial Position	Length of List							
	8	9	10	11	12	13	14	15
1	22.2	20.0	18.2	16.5	15.0	14.1	13.3	12.6
2	19.4	17.7	16.4	15.0	13.8	13.0	12.4	11.0
3	12.5	11.7	11.4	10.9	10.5	10.0	9.6	9.3
4	9.7	9.3	9.5	9.3	9.2	8.9	8.7	8.6
5	6.9	7.4	6.8	6.5	6.2	6.2	6.3	6.2
6	6.9	5.5	5.0	5.0	5.0	5.2	5.3	5.2
7	9.7	7.4	5.0	4.8	4.7	4.3	4.1	4.0
8	12.5	9.3	6.8	5.0	4.7	3.8	3.1	3.5
9		11.7	9.5	6.5	5.0	4.3	3.1	2.9
10			11.4	9.3	6.2	5.2	4.1	3.5
11				10.9	9.2	6.2	5.3	4.0
12					10.5	8.9	6.3	5.2
13						10.0	8.7	6.2
14							9.6	8.6
15								9.3

Note.—Entries are percentages of total correct responses for list of each length.

points," selecting the next items to be learned at random from either anchor point. The manner in which we obtain the numerical predictions in Table 1 is spelled out below. A much more detailed account of the process is available in two previous reports (Feigenbaum, 1959; Feigenbaum & Simon, 1958).

The numerical predictions in Table 1 were obtained as follows: We assume the syllables will be learned in an orderly sequence, each syllable requiring a certain processing time, say k. Each syllable (s) can be identified by its serial order, i, in the list as presented by the experimenter, and also by the order, r, in which it is learned by the subject. Since learning takes place from both ends of the list, these two orders will not, in general, be identical. Thus, s_i, the ith syllable in order of presentation, may be the same syllable as s_r', the rth syllable in order of learning. (Technically, the list of syllables in order of learning is a permutation of the original list.)

Let T_r' be the time that elapses before the first successful response to syllable s_r'—that is, until the rth syllable is learned. Then $T' = kr$, and the number of errors, W_r', the subject will make on the rth syllable, which is equal to the number of learning trials prior to the trial on which that syllable is learned, will be proportional to r:

$$W_r' = mr \qquad [1]$$

where m is a proportionality constant, equal to k divided by the time per trial. The numerical value of m is a function, of course, of the difficulty of the item, and the rate of presentation.

If the list is of length L, we will then have $W_L' = mL$; and if the list is learned to a criterion of one correct response, we will have for N, the number of trials to criterion:

$$N = W_L' + 1 = mL + 1 \qquad [2]$$

Then the number of correct responses, C_r', the subject will make on the rth syllable, is:

$$C_r' = N - W_r' = mL + 1 - mr$$
$$= 1 + m(L - r) \qquad [3]$$

Finally, c_r', the *percentage* of total correct responses made on the rth syllable,

will be:

$$c_r' = C_r'/\Sigma C_r'$$

$$= \frac{1 + m(L - r)}{L + m(L^2 - \sum_1 r)}$$

$$= \frac{1 + m(L - r)}{L + m\left(L^2 - \frac{L(L+1)}{2}\right)}$$

$$= \frac{1/m + (L - r)}{L/m + \frac{1}{2}(L^2 - L)} \quad [4]$$

It will be seen from Equation 4 that c_r' depends only weakly on m. In our calculations, we have assumed $m = 1$.

To apply this result we must calculate the rank, r_i, of the ith syllable in order of presentation, as determined by our theory.

We assume that the immediate memory capacity is two syllables (the results do not depend in any sensitive way on this assumption). Then the first two syllables in the list will be learned first (will have rank, $r = 1$ and $r = 2$, respectively), followed either by the last two syllables on the list, or by Syllables 3 and 4, each with probability ½. Then, subsequent pairs will be chosen from the beginning and end of the list of those not yet learned, always with probability ½. The result is that in a list of 12 syllables the third syllable, for example, will have a probability of ½ of being the third syllable in order of learning, a probability of ¼ of being the fifth syllable, a probability of ⅛ of being seventh, and of 1/16 of being ninth or eleventh. Averaging these ranks, weighted by their respective probabilities, we find that the average rank of the third syllable is $r_3 = 4.875$.

If we now substitute $r_3 = 4.875$, $m = 1$, and $L = 12$ in Equation 4, we obtain $c_3' = .105$, the value in Table 1 for list length 12. The other values in the table are obtained in the same way. To see how insensitive the result is to the value of m, we observe that for $m = 2$, the same calculation gives $c_3' = .106$.

In Table 1, we use our theory to make predictions that may be compared with those of Table 5 in Murdock's paper.

TABLE 2

PREDICTED AND OBTAINED RESULTS FROM BUGELSKI (1950)

Serial Position	Predicted (Murdock)	Predicted (F & S)	Actual (Bugelski)
1	22.9	22.2	23.8
2	14.0	19.4	18.1
3	10.4	12.5	12.3
4	9.2	9.7	9.2
5	9.2	6.9	7.9
6	10.0	6.9	6.7
7	11.4	9.7	8.9
8	12.9	12.5	13.1

Note.—Entries in table are percentages.

It should be possible, with empirical data, to choose between the two theories, because the predictions differ systematically in several respects. In general, our theory predicts more bowing for long lists than does Murdock's. We invite experimenters, as does Murdock, to generate some data that can be used to test these predictions.

In Table 2, we compare our predictions for a list of eight syllables with the predictions of Murdock's theory and with the empirical data he provides in his Table 6. In the case of five of the eight serial positions, our prediction is better than his; in the case of three positions, his is better. The mean absolute error of his predictions is 1.8 percentage points, while the mean absolute error of our predictions is only .8 percentage points. His maximum error is 4.1 percentage points; ours is 1.6 percentage points. Thus our maximum error is less than his average error, and about equal to his median error. In this single test, we conclude that our theory predicts the data better than the theory based on distinctiveness.

REFERENCES

ATKINSON, R. C. A stochastic model for rote serial learning. *Psychometrika,* 1957, **22,** 87-95.

BUGELSKI, B. R. A remote association explanation of the relative difficulty of learning nonsense syllables in a serial list. *J. exp. Psychol.,* 1950, **40,** 336-348.

Feigenbaum, E. A. *An information processing theory of verbal learning.* (Report P-1817) Santa Monica: RAND, 1959.

Feigenbaum, E. A., & Simon, H. A. *A theory of the serial position effect.* (CIP Working Paper No. 14) Pittsburgh: Carnegie Institute of Technology, 1958. (Ditto)

Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.

Murdock, B. B., Jr. The distinctiveness of stimuli. *Psychol. Rev.*, 1960, **67**, 16–31.

(Received February 11, 1960)

Psychological Review
1961, Vol. 68, No. 4, 288–292

A BIOCHEMICAL APPROACH TO LEARNING AND MEMORY

JOHN GAITO[1]

Lake Forest College

The problem of learning and memory has been one of the most important areas of investigation for the psychologist for many years. Recently there have been a number of attempts to specify the neurological structure which is modified during learning and which provides the basis for memory. It appears that the site which has been emphasized most frequently by theorists is the synapse. Hebb (1949) maintained that reverberatory circuits may provide the basis for short term memory whereas long term memory requires neural growth at synaptic junctions; his cell assemblies and phase sequences furnish the basis for all simple and complex functions. Somewhat similar ideas (stripped of cell assemblies and phase sequences) were expressed by a number of individuals at the Hixon Symposium (Jeffress, 1951) and at the Laurentian Symposium (Delayfresnaye, 1954). Gerard (1953) mentioned a number of events which he believed might provide the synaptic changes which facilitate learning and maintain memory, viz., swelling of synaptic nerve fiber end-bulbs while conducting impulses, alterations of potential, actual neural growth at synapse, and changes in nerve proteins. Eccles (1958) presented a somewhat similar view. Krech and his collaborators at the University of California have conducted a series of experiments combining the biochemical and psychological approaches and have emphasized that experience and training may significantly alter the concentration of brain cholinesterase (Bennett, Krech, Rosenzweig, Karlsson, Dye, & Ohlander, 1958; Bennett, Rosenzweig, Krech, Karlsson, Dye, & Ohlander, 1958; Krech, Rosenzweig, & Bennett, 1956, 1959; Krech, Rosenzweig, Bennett, & Longueil, 1959; Rosenzweig, Krech, & Bennett, 1956). These individuals have hypothesized that transmission of neural impulses is accomplished by the discharge of actylcholine from presynaptic neurons. Inactivation of acetylcholine by cholinesterase (as soon as the former stimulates the postsynaptic neuron) preserves discrete transmission of impulses. Their experimental work lends credence to this hypothesis. Overton (1958, 1959a, 1959b) stated that calcium displacement at the synapse, during acetycholine-cholinesterase activity, forms the basis for memory. Likewise, he reported research which favors his hypothesis.

That some synaptic change is important for learning and memory is generally accepted. However, it is possible that these synaptic changes *facilitate or set off reactions* which allow changes to occur elsewhere in the nerve cell. Recently Halstead (1951) favored the hypothesis that nucleoproteins are the substances which have the ability to act as templates on which replica molecules are formed,

[1] Now at Kansas State University.

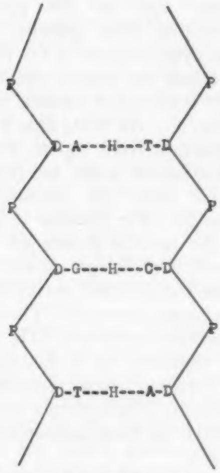

FIG. 1. The DNA molecule as it would appear if unwound from its helical formation. (P is phosphate; D, desoxyribose, a sugar; A, adenine; T, thymine; C, cytosine; G, guanine; and H, hydrogen.)

both genetically and as a result of individual experience. He believed that templates exist in the brain nerve cells which are like those of the germ cells in representing native endowment but differ from the latter in having templates arising from external stimulation. He stated that the ordering of the protein templates could take place in various components of the cell and its processes. The present arguments are in general similar to those of Halstead but emphasize recent biochemical developments. We suggest that *learning and memory depend on changes in genic material (or the by-products of genic activity) either in the nucleus or cytoplasm of the nerve cell soma.*

It is obvious to psychologists that both heredity and environment contribute to behavior. Thus we can say, in very general fashion, that behavior is a function of genetic potential (heredity) as it is modified by intra- and extraorganismic environmental stimulation. Inasmuch as we are concerned only with learning (and resulting memory), let us substitute the word "learning" for the second part of the above statement. We now have: behavior is a function of genetic potential as it is modified by learning. Also we shall take this statement literally.

First, let us indicate the physical basis for genetic potential. Biochemists (e.g., Butler, 1959; Crick, 1954, 1957; Lederberg, 1960) have identified desoxyribonucleic acid (DNA), a giant molecule found in the chromosomes, as the primary hereditary material. DNA is a large double strand molecule which is wound in a helix. Each strand is a mirror image of the other and has recurring patterns of constituents throughout its length, called nucleotides. A nucleotide consists of a phosphate attached to a sugar-base linkage. Figure 1 shows a portion of the DNA molecule as it would appear if it were unwound. The sugar-base portion (a nucleoside) is considered to be the most important part of the molecule. The nucleoside of one strand is attached to its corresponding part of the other strand by a hydrogen atom. The bases consist of two types: purines and pyrimidines. Furthermore, there are two purines (adenine and guanine) and two pyrimidines (thymine and cytosine). The purines are larger molecules than are the pyrimidines. At the sugar-base points of the strands, a purine of one strand is always attached (by the hydrogen atom) to a pyrimidine. Two purines are too big to bridge the gap between the two strands and two pyrimidines are too small. Furthermore, the amounts of adenine and thymine are always equal and the amounts of guanine and cytosine are equal also, because adenine is always paired with thymine and guanine with cytosine. Thus there are two basic types. However, if we consider the order of the bases in the attachment of one nucleoside to another, this number doubles. (For example, adenine linked to thymine is different than thymine linked to adenine.) Thus there are four basic types of nucleotides.

Biochemists believe that the sequence of the bases furnish the basis for the "codes" of genetic potential (Butler, 1959; Crick, 1954, 1957; Lederberg, 1960). Even though there are only four possible nucleotides, the strands of the

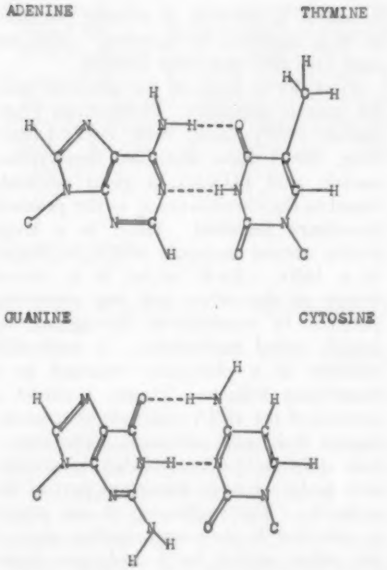

FIG. 2. The bases present in DNA. (RNA is similar except that uracil replaces thymine. The point of bonding is indicated by dashed lines. C is carbon; N, nitrogen; H, hydrogen; and O, oxygen.)

molecule are extremely long and allow for infinite possibilities, if the sequence of the purines and pyrimidines is considered. Thus, these sugar-base attachments may be the "language" of the genes. Crick (1954) has likened the bases to the dots and dashes of the Morse Code. Thus he states that there is enough DNA in a single cell of the human body to encode about 1,000 large textbooks.

Now let us see how learning might modify these codes of the DNA molecule. DNA is a relatively stable molecule. However, changes of one or several nucleosides should be easier to effect. Thus electrochemical changes at the synapse might spread to the soma and lead to the dislocation or changing of nucleosides at one or more loci. A change at one locus would modify the code.

Figure 2 indicates the structure of the purines and pyrimidines. Note that the basic structure of the purines is the same, as is also that of the pyrimidines. Adenine differs from guanine and thymine from cytosine only in the side bonds. It appears that the easiest and most logical change would be of adenine to guanine or the reverse. An NH_2 side bond must change position from top to bottom and a hydrogen atom must be replaced by an oxygen atom for adenine to become guanine. For guanine to change to adenine, the reverse events are required.

However, if a change were to occur in one purine, the attached pyrimidine would have to change likewise. Thus for thymine to become cytosine a CH_3 side chain must be replaced by a hydrogen atom and an oxygen atom replaced by an NH_2 molecule. Such events appear feasible but have not been indicated biochemically as yet.

Another possibility is that changes which represent learning and memory involve ribonucleic acid (RNA). The structure of the RNA molecule is not known completely; however, it is somewhat similar to DNA but contains uracil rather than thymine and ribose sugar rather than desoxyribose sugar. RNA is found in the nucleus and cytoplasm of cells. It is believed that RNA carries "instructions" from the genes (through operation of the DNA molecule) which allows it to direct the assembly of proteins (Hoagland, 1959). The suggested modifications of the RNA molecules through learning would be similar to those hypothesized above for the DNA and would involve the bases with one purine changing to the other and uracil becoming cytosine, or vice versa.

Another possible cytoplasmic or nucleic molecule upon which learning might impose its effects is that of the proteins. Like DNA and RNA, the proteins are large molecules and consist of long chains with small molecules, the amino acids, of which there are 20 different kinds. A protein molecule may contain as many as 100 amino acid units. The learning effect would involve some modification of one or more amino acids. That an amino acid change can greatly modify behavior has been indicated recently. Ingram (1958) has reported that the

change of a single amino acid in hemoglobin is responsible for the appearance of sickle cell anemia. He found that the normal hemoglobin cell and the sickle form hemoglobin are the same except that in the latter, one peptide was displaced slightly from the position it occupies in the normal cell. The normal peptide had two glutamic acid units and one valine unit; the abnormal cell had one glutamic acid unit and two valines. This defect is due to heredity or to mutations. However, changes in amino acids might occur through stimulation.

Thus we believe that changes in the DNA, RNA, or protein molecules may be the basis for learning and memory. These three are intimately involved in cellular activity in the nucleus and cytoplasm with the DNA commanding the process. It is believed that chromosomes consist of these three, with the DNA being the backbone of the chromosome and housing the genes. If changes occur in the DNA molecules, these changes would automatically affect RNA and protein synthesis. Likewise, changes in RNA would be transmitted to some protein molecules. However, changes might occur to protein molecules, which would have no effect on DNA or RNA. Only the changes of the DNA molecule would be of a genic modification nature. Changes of DNA, RNA, or protein molecules might occur in the nucleus or in the cytoplasm. Gay (1960) has described the passage of minute "blebs," or blisters, containing chromosomal material, from the nucleus to the cytoplasm in the salivary glands of the larval fruit fly. These events are considered of importance in protein synthesis. It is possible that such transfer could occur also in nerve cells. Thus "recodification" of the DNA, RNA, or protein molecules might occur in the cytoplasm following such transfer. However, it should be possible for electrochemical events to permeate the nuclear membrane and bring about the changes in the nucleus.

Superficially there appears to be conflict between the synaptic hypothesis and the ideas presented here. However, this need not be the case. The synaptic hypothesis allows for changes to occur elsewhere in nerve cells but concentrates on the synaptic changes in effecting learning and memory. On the other hand the present viewpoint agrees that changes occur at the synapse but stresses the modification in the genic, or by-products of the genic, material; the synaptic changes are preliminary in nature. Furthermore, it is probable that both mechanisms are involved in learning and memory. In recent years there has been a tendency to consider memory as of two types, transient and permanent (e.g., Hebb, 1949; Jeffress, 1951), with the former maintained by reverberatory circuits but the latter, by permanent neural changes. It is possible that the two types of memory may depend upon different neurological sites. For example, a reverberatory circuit would depend upon synaptic transmission for reverberation to occur. Thus effecting short term memory may involve mainly the synapse. However, the neural change which makes for relatively permanent memory may involve changes elsewhere in the cell, e.g., in the DNA, RNA, or amino acids in the nucleus or in the cytoplasm. But permanent changes could, and apparently do, occur at the synapse. However, we believe that the synaptic changes serve the function of making the genic materials more accessible to stimulation.

The major advantage that this nucleic or cytoplasmic recodification hypothesis has over the synaptic hypothesis is that the former allows learning to directly modify genetic potential (or the by-products of genic activity) whereas the latter is faced with the problem of relating learning and memory changes to the genetic potential. Thus the synaptic hypothesis must explain the means by which learning and genetic potential affect one another, for it is obvious that genetic potential affects learning and that learning modifies genetic potential. Furthermore, the recodification hypothesis attempts to indicate the molecular basis for memory; however, the exact chemical mechanism involved in the learning process is still a major problem.

One might criticize portions of the

above hypothesis on the ground that it is arguing for the inheritance of acquired characteristics. However, this is not so. The suggested changes in genic material would occur only in nervous tissue which is not passed on to the next generation. The genes in the nuclear chromosomes of the gonads (which provide the physical material for heredity) would be unaffected.

The validity of these hypotheses have not been demonstrated as yet. To evaluate such ideas (or similar ones) requires the use of the precise molecular techniques of the biochemist, cytologist, and geneticist coupled with the less molecular procedures of the neurophysiologist and the molar learning methods of the psychologist in collaborative research programs. However, the hypothesis can serve as an interesting framework for descriptive formulations and appears more logical than does the synaptic hypothesis.

REFERENCES

BENNETT, E. L., KRECH, D., ROSENZWEIG, M. R., KARLSSON, H., DYE, N., & OHLANDER, A. Cholinesterase and lactic dehydrogenase activity in the rat brain. *J. Neurochem.*, 1958, **3**, 153–160.

BENNETT, E. L., ROSENZWEIG, M. R., KRECH, D. R., KARLSSON, H., DYE, N., & OHLANDER, A. Individual, strain and age differences in cholinesterase activity of the rat brain. *J. Neurochem.*, 1958, **3**, 144–152.

BUTLER, J. A. V. *Inside the living cell.* New York: Basic Books, 1959.

CRICK, F. H. C. The structure of the hereditary material. *Scient. Amer.*, 1954, **191**, 54–61.

CRICK, F. H. C. Nucleic acids. *Scient. Amer.*, 1957, **197**, 188–200.

DELAYFRESNAYE, J. F. (Ed.) *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954.

ECCLES, J. C. The physiology of imagination. *Scient. Amer.*, 1958, **199**, 135–146.

GAY, H. Nuclear control of the cell. *Scient Amer.*, 1960, **202**, 126–136.

GERARD, R. W. What is memory? *Scient. Amer.*, 1953, **189**, 118–126.

HALSTEAD, W. C. Brain and intelligence. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HOAGLAND, M. B. Nucleic acids and proteins. *Scient. Amer.*, 1959, **201**, 55–61.

INGRAM, V. M. How do genes act? *Scient. Amer.*, 1958, **198**, 68–74.

JEFFRESS, L. A. (Ed.) *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

KRECH, D., ROSENZWEIG, M. R., & BENNETT, E. L. Dimensions of discrimination and level of cholinesterase activity in the cerebral cortex of the rat. *J. comp. physiol. Psychol.*, 1956, **49**, 261–268.

KRECH, D., ROSENZWEIG, M. R., & BENNETT, E. L. Correlation between brain cholinesterase and brain weight within two strains of rats. *Amer. J. Physiol.*, 1959, **196**, 31–32.

KRECH, D., ROSENZWEIG, M. R., BENNETT, E. L., & LONGUEIL, C. L. Changes in brain chemistry of the rat following experience. Paper presented at American Psychological Association, Cincinnati, September 1959.

LEDERBERG, J. A view of genetics. *Science,* 1960, **131**, 269–276.

OVERTON, R. K. An effect of high- and low-calcium diets on the maze performance of rats. *J. comp. physiol. Psychol.*, 1958, **51**, 697–699.

OVERTON, R. K. The calcium displacement hypothesis: A review. *Psychol. Rep.*, 1959, **5**, 721–724. (a)

OVERTON, R. K. *Thought and action: A physiological approach.* New York: Random House, 1959. (b)

ROSENZWEIG, M. R., KRECH, D., & BENNETT, E. L. Effects of pentobarbital sodium on adaptive behavior patterns in the rat. *Science,* 1956, **123**, 371–372.

(Received March 10, 1960)

PSYCHOLOGICAL MONOGRAPHS

Connotative Meaning as a Determinant of Stimulus Generalization, by CHARLES F. DICKEN

An experimental study of a set of three hypotheses derived from Osgood's approach to the nature and measurement of meaning

No. 505. Price $1.50

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington 6, D. C.

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by BORIS SEMEONOFF

Vol. 52 Part 2 May, 1961 20s. net

The subscription price per volume, payable in advance, is 60s. net (post free).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

Bentley House, Euston Road, London, N. W. 1